Helpful microbes take root down on the farm *p. 680*

The so-called global warming hiatus revisited *p. 691*

A promising Ebola virus vaccine *pp. 693 & 739*

Science

$10
14 AUGUST 2015
sciencemag.org

AAAS

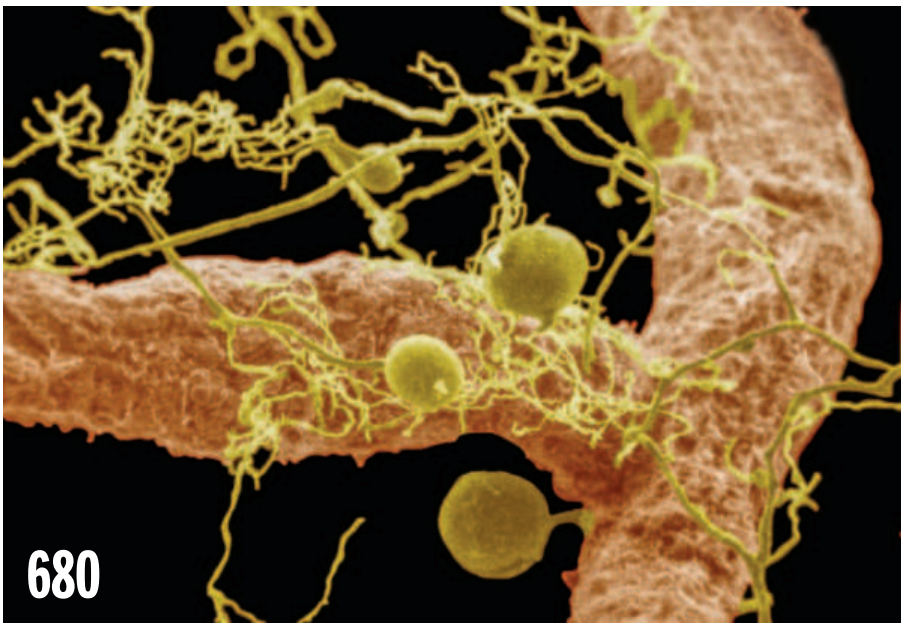
Giving voice

Early vocal development in primates *pp. 688 & 734*

CONTENTS

704
The future of carbon transistors

14 AUGUST 2015 • VOLUME 349 • ISSUE 6249



680

NEWS

IN BRIEF

672 Roundup of the week's news

IN DEPTH

674 **IRAN'S ATOMIC CZAR DESCRIBES THE ART OF THE DEAL**
Ali Akbar Salehi tells *Science* how he helped seal the Iran nuclear agreement *By R. Stone*

676 **REPORT SETS NEW GOALS FOR U.S. ANTARCTIC PROGRAM**
Committee pushes study of ice loss, genomics, and radiation from the beginning of the universe *By C. Gramling*

677 **MODIFIED YEAST PRODUCE OPIATES FROM SUGAR**
Engineered microbes raise hopes for better medicines and concerns about illicit drugs *By R. F. Service*
▸ REPORT BY S. GALANIE *ET AL.* 10.1126/science.aac9373

678 **RUSSIAN CLAIM HEATS UP BATTLE TO CONTROL ARCTIC SEA FLOOR**
Nations offer new geological data to back overlapping bids to expand marine territory claims *By L. Laursen*

679 **MASHCO PIRO TRIBE EMERGES FROM ISOLATION IN PERU**
Crisis sparks debate about whether to make official contact or leave isolated tribespeople alone *By A. Lawler*

FEATURE

680 **THE LITTLEST FARMHANDS**
Scientists are discovering thousands of microbes that help plants survive and thrive. Could these symbionts help farmers as well? *By J. de Vrieze*

INSIGHTS

PERSPECTIVES

WATCHING WATER: FROM SKY OR STREAM?

684 **Satellites provide the big picture**
By J. S. Famiglietti et al.

685 **Time for in situ renaissance**
By B. M. Fekete et al.

687 **HOW MOUNTAINS GET MADE**
Observations of crustal deformation constrain models of mountain formation *By M. D. Long*
▸ REPORT P. 720

688 **MARMOSET KIDS ACTUALLY LISTEN**
Humans and other vertebrates may share a developmental program for vocal learning
By D. Margoliash and O. Tchernichovski
▸ REPORT P. 734

690 **ASTROCYTES TELL NEURONS WHEN TO LISTEN UP**
Astrocytes ensure communication between neurons of the same circuit
By A. H. Gittis and D. J. Brasier
▸ REPORT P. 730

691 **HAS THERE BEEN A HIATUS?**
Internal climate variability masks climate-warming trends
By K. E. Trenberth

693 **EBOLA VIRUS VACCINES—PREPARING FOR THE UNEXPECTED**
Many lives might have been saved if clinical studies of Ebola virus vaccines had been done earlier
By H.-D. Klenk and S. Becker
▸ REPORT P. 739

694 **TUMOR ANGIOGENESIS, FROM FOE TO FRIEND**
Treatments that normalize or even promote blood vessel growth may enhance drug delivery to tumors
By L. B. Rivera and G. Bergers

696 **THE ON-RAMP TO THE ALL-OPTICAL QUANTUM INFORMATION PROCESSING HIGHWAY**
An optical waveguide circuit can be flexibly programmed with near-perfect fidelity *By P. P. Rohde and J. P. Dowling*
▸ RESEARCH ARTICLE P. 711

BOOKS *ET AL.*

697 **GODS OF THE MORNING**
By J. Lister-Kaye, reviewed by C. Kemp

697 **THE MATTER FACTORY**
By P. J. T. Morris

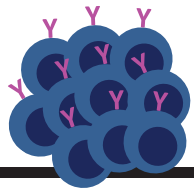
698 **THE VILLAGE EFFECT**
By S. Pinker, reviewed by B. Wellman

LETTERS

699 **PARKS FOR SCIENCE, SCIENCE FOR PARKS**
By W. F. Laurance

Published by AAAS

687 & 720
Whence the mountains in Taiwan?



705
Microbiota steer an HIV vaccine off target

699 **DISASTER RESPONSE: VOLUNTEERS, UNITE!**
By U. B. Shrestha et al.

699 **ONLINE BUZZ: CAREERS**

RESEARCH

IN BRIEF

701 From *Science* and other journals

REVIEW

704 **DEVICE TECHNOLOGY**
Nanomaterials in transistors: From high-performance to thin-film applications *A. D. Franklin*
REVIEW SUMMARY; FOR FULL TEXT: dx.doi.org/10.1126/science.aab2750

RESEARCH ARTICLES

705 **HIV-1 VACCINES**
Diversion of HIV-1 vaccine–induced immunity by gp41-microbiota cross-reactive antibodies *W. B. Williams* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: dx.doi.org/10.1126/science.aab1253

706 **PALEOCEANOGRAPHY**
A warm and poorly ventilated deep Arctic Mediterranean during the last glacial period *D. J. R. Thornalley* et al.

711 **QUANTUM OPTICS**
Universal linear optics *J. Carolan* et al.
▸ PERSPECTIVE P. 696

REPORTS

716 **MAGNETISM**
Imaging and control of ferromagnetism in $LaMnO_3/SrTiO_3$ heterostructures *X. R. Wang* et al.

720 **GEOPHYSICS**
Layered deformation in the Taiwan orogen *T.-Y. Huang* et al.
▸ PERSPECTIVE P. 687

723 **2D MATERIALS**
Observation of tunable band gap and anisotropic Dirac semimetal state in black phosphorus *J. Kim* et al.

726 **APPLIED OPTICS**
Voltage-tunable circular photogalvanic effect in silicon nanowires *S. Dhara* et al.

730 **GLIAL CELL SIGNALING**
Circuit-specific signaling in astrocyte-neuron networks in basal ganglia pathways *R. Martín* et al.
▸ PERSPECTIVE P. 690

734 **LANGUAGE DEVELOPMENT**
The developmental dynamics of marmoset monkey vocal production *D. Y. Takahashi* et al.
▸ PERSPECTIVE P. 688; PODCAST

739 **EBOLA VACCINE**
VSV-EBOV rapidly protects macaques against infection with the 2014/15 Ebola virus outbreak strain *A. Marzi* et al.
▸ PERSPECTIVE P. 693

742 **DNA REPAIR**
Mus81 and converging forks limit the mutagenicity of replication fork breakage *R. Mayle* et al.

747 **EVOLUTION**
Fruit flies diversify their offspring in response to parasite infection *N. D. Singh* et al.

DEPARTMENTS

671 **EDITORIAL**
Science, all inclusive
By Shirley M. Malcom

758 **WORKING LIFE**
Creating a research brand
By Jeffrey J. McDonnell

ON THE COVER

Infant common marmoset (*Callithrix jacchus*) in the act of vocalizing. For decades, it has been assumed that monkeys are born with adult-sounding vocalizations. However, Takahashi *et al.* report that marmoset infant vocalizations, like those of human babies, change substantially in early life and are influenced by parental interaction. See pages 688 and 734.
Photo: Emmanuel Keller

Science Staff 670
New Products 751
Science Careers 752

PHOTO: (FLY) DAHLIA NIELSEN/NC STATE UNIVERSITY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2015 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $153 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1282; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $30.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Bateman House, 82-88 Hills Road, Cambridge, UK CB2 1LQ

Editor-in-Chief Marcia McNutt
Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller
Managing Editor, Research Journals Katrina L. Kelner
Deputy Editors Barbara R. Jasny, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

Research and Insights

SR. EDITORS Caroline Ash(UK), Gilbert J. Chin, Lisa D. Chong, Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurtley(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Kristen L. Mueller, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuromi, Brad Wible, Nicholas S. Wigginton, Laura M. Zahn **ASSOCIATE EDITORS** Brent Grocholski, Keith T. Smith, Sacha Vignieri **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **ASSOCIATE LETTERS EDITOR** Jennifer Sills **CHIEF CONTENT PRODUCTION EDITOR** Cara Tate **SR. CONTENT PRODUCTION EDITOR** Harry Jach **CONTENT PRODUCTION EDITORS** Jeffrey E. Cook, Chris Filiatreau, Cynthia Howe, Lauren Kmec, Barbara P. Ordway, Catherine Wolner **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Ramatoulaye Diop, Joi S. Granger, Lisa Johnson, Anita Wynn **PUBLICATIONS ASSISTANTS** Aneera Dobbins, Jeffrey Hearn, Dona Mathieu, Le-Toya Mayne Flood, Shannon McMahon, Scott Miller, Jerry Richardson, Rachel Roberts(UK), Alice Whaley(UK), Brian White **EXECUTIVE ASSISTANT** Anna Bashkirova **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Lizanne Newton(UK), Maryrose Madrid, Laura-Nadine Schuhmacher (UK, Intern), Alix Welch (Intern), John Wood(UK)

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Richard Stone **DEPUTY NEWS EDITORS** Daniel Clery(UK), Robert Coontz, Elizabeth Culotta, David Grimm, David Malakoff, Leslie Roberts **CONTRIBUING EDITOR** Martin Enserink(Europe) **SR. CORRESPONDENTS** Jeffrey Mervis, Elizabeth Pennisi **NEWS WRITERS** Adrian Cho, Jon Cohen, Jennifer Couzin-Frankel, Carolyn Gramling, Eric Hand, Jocelyn Kaiser, Catherine Matacic, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Emily Underwood **INTERNS** Hanae Armitage, Emily DeMarco, Annick Laurent, Laura Olivieri, Juan David Romero **CONTRIBUTING CORRESPONDENTS** Michael Balter(Paris), John Bohannon, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Christina Larson(Beijing), Mitch Leslie, Charles C. Mann, Eliot Marshall, Virginia Morell, Dennis Normile(Tokyo), Heather Pringle, Tania Rabesandratana(London), Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein **COPY EDITORS** Julia Cole, Jennifer Levin (Chief) **ADMINISTRATIVE SUPPORT** Jessica Williams

Executive Publisher Rush D. Holt
Publisher Kent R. Anderson **Chief Digital Media Officer** Rob Covey

BUSINESS OPERATIONS AND PORTFOLIO MANAGEMENT DIRECTOR Sarah Whalen **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS DIRECTOR** Randy Yi **MANAGER OF FULFILLMENT SYSTEMS** Neal Hawkins **SYSTEMS ANALYST** Nicole Mehmedovich **ASSISTANT DIRECTOR, BUSINESS OPERATIONS** Eric Knott **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **BUSINESS ANALYSTS** Cory Lipman, Cooper Tilton, Celeste Troxler **FINANCIAL ANALYST** Robert Clark **RIGHTS AND PERMISSIONS ASSISTANT DIRECTOR** Emilie David **PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Kiser

MARKETING DIRECTOR Ian King **MARKETING MANAGER** Julianne Wielga **MARKETING ASSOCIATE** Elizabeth Sattler **SR. MARKETING EXECUTIVE** Jennifer Reeves **SR. ART ASSOCIATE, PROJECT MANAGER** Tzeitel Sorrosa **ART ASSOCIATE** Seil Lee **JR. ART ASSOCIATE** Kim Huynh **ASSISTANT COMMERCIAL EDITOR** Selby Frame **MARKETING PROJECT MANAGER** Angelissa McArthur **PROGRAM DIRECTOR, AAAS MEMBER CENTRAL** Peggy Mihelich **FULFILLMENT SYSTEMS AND OPERATIONS** membership@aaas.org **MANAGER, MEMBER SERVICES** Pat Butler **SPECIALISTS** LaToya Casteel, Terrance Morrison, Latasha Russell **MANAGER, DATA ENTRY** Mickie Napoleoni **DATA ENTRY SPECIALISTS** JJ Regan, Brenden Aquilino, Fiona Giblin

DIRECTOR, SITE LICENSING Tom Ryan **DIRECTOR, CORPORATE RELATIONS** Eileen Bernadette Moran **SR. PUBLISHER RELATIONS SPECIALIST** Kiki Forsythe **PUBLISHER RELATIONS MANAGER** Catherine Holland **PUBLISHER RELATIONS, EASTERN REGION** Keith Layson **PUBLISHER RELATIONS, WESTERN REGION** Ryan Rexroth **SALES RESEARCH COORDINATOR** Aiesha Marshall **MANAGER, SITE LICENSE OPERATIONS** Iquo Edim **SENIOR PRODUCTION SPECIALIST** Robert Koepke **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT ANALYST** Judy Lillibridge **ASSOCIATE DIRECTOR, MARKETING** Christina Schlecht **MARKETING ASSOCIATES** Thomas Landreth, Isa Sesay-Bah

DIRECTOR OF WEB TECHNOLOGIES Ahmed Khadr **SR. DEVELOPER** Chris Coleman **DEVELOPERS** Dan Berger, Jimmy Marks **SR. PROJECT MANAGER** Trista Smith **SYSTEMS ENGINEER** Luke Johnson

CREATIVE DIRECTOR, MULTIMEDIA Martyn Green **DIRECTOR OF ANALYTICS** Enrique Gonzales **SR. WEB PRODUCER** Sarah Crespi **WEB PRODUCER** Alison Crawford **VIDEO PRODUCER** Nguyen Nguyen **SOCIAL MEDIA PRODUCER** Meghna Sachdev

DIRECTOR OF OPERATIONS PRINT AND ONLINE Lizabeth Harman **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **PROJECT ACCOUNT MANAGER** Tara Kelly **DIGITAL PRODUCTION MANAGER** Lisa Stanford **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy, Anthony Rosen **CONTENT SPECIALISTS** Jacob Hedrick, Kimberley Oster

DESIGN DIRECTOR Beth Rakouskas **DESIGN EDITOR** Marcy Atarod **SENIOR DESIGNER** Garvin Grullón **DESIGNER** Chrystal Smith **GRAPHICS MANAGING EDITOR** Alberto Cuadra **SENIOR SCIENTIFIC ILLUSTRATORS** Chris Bickel, Katharine Sutliff **SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SENIOR ART ASSOCIATES** Holly Bishop, Preston Huey **SENIOR PHOTO EDITOR** William Douthitt **PHOTO EDITORS** Leslie Blizard, Christy Steele

DIRECTOR, GLOBAL COLLABORATION, CUSTOM PUBLICATIONS, ADVERTISING Bill Moran **EDITOR, CUSTOM PUBLISHING** Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Tianna Hicklin: 202-326-6463 **ADVERTISING MARKETING MANAGER** Justin Sawyers: 202-326-7061 science_advertising@aaas.org **ADVERTISING MARKETING ASSOCIATE** Javia Flemmings **ADVERTISING SUPPORT MANAGER** Karen Foote: 202-326-6740 **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **PRODUCTION SPECIALIST** Yuse Lajiminmuhip **SR. TRAFFIC ASSOCIATE** Christine Hall **SALES COORDINATOR** Shirley Young **ASSOCIATE DIRECTOR, COLLABORATION, CUSTOM PUBLICATIONS/CHINA/TAIWAN/KOREA/SINGAPORE** Ruolei Wu: +86-186 0822 9345, rwu@aaas.org **COLLABORATION/ CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Katsuyoshi Fukamizu(Tokyo): +81-3-3219-5777 kfukamizu@aaas.org **CHINA/TAIWAN** Ruolei Wu: +86-186 0822 9345, rwu@aaas.org

WORLDWIDE ASSOCIATE DIRECTOR OF SCIENCE CAREERS Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532 tholmes@science-int.co.uk **CLASSIFIED** advertise@sciencecareers.org **U.S. SALES** Tina Burks: 202-326-6577, Nancy Toema: 202-326-6578 **SALES ADMINISTRATOR** Marci Gallun **EUROPE/ROW SALES** Axel Gesatzki, Sarah Lelarge **SALES ASSISTANT** Kelly Grace **JAPAN** Hiroyuki Mashiki(Kyoto): +81-75-823-1109 hmashiki@aaas.org **CHINA/TAIWAN** Ruolei Wu: +86-186 0082 9345 rwu@aaas.org **MARKETING MANAGER** Allison Pritchard **MARKETING ASSOCIATE** Aimee Aponte

AAAS BOARD OF DIRECTORS RETIRING PRESIDENT, CHAIR Gerald R. Fink **PRESIDENT** Geraldine (Geri) Richmond **PRESIDENT-ELECT** Barbara A. Schaal **TREASURER** David Evans Shaw **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Bonnie L. Bassler, May R. Berenbaum, Carlos J. Bustamante, Stephen P.A. Fodor, Claire M. Fraser, Michael S. Gazzaniga, Laura H. Greene, Elizabeth Loftus, Mercedes Pascual

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6755 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Services** 202-326-6417 or http://membercentral.aaas.org/discounts

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority of conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and books reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

INFORMATION FOR AUTHORS See pages 678 and 679 of the 6 February 2015 issue or access www.sciencemag.org/about/authors

SENIOR EDITORIAL BOARD

Robert H. Grubbs, *California Institute of Technology,* **Gary King,** *Harvard University*
Susan M. Rosenberg, *Baylor College of Medicine,* **Ali Shilatifard,** *Northwestern University*
Feinberg School of Medicine, **Michael S. Turner**, *U. of Chicago*

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)

Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Judith Allen, *U. of Edinburgh*
Sonia Altizer, *U. of Georgia*
Sebastian Amigorena, *Institut Curie*
Kathryn Anderson, *Memorial Sloan-Kettering Cancer Center*
Meinrat O. Andreae, *Max-Planck Inst. Mainz*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Jordi Bascompte, *Estación Biológica de Doñana CSIC*
Facundo Batista, *London Research Inst.*
Ray H. Baughman, *U. of Texas, Dallas*
David Baum, *U. of Wisconsin*
Carlo Beenakker, *Leiden U.*
Kamran Behnia, *ESPCI-ParisTech*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Stephen J. Benkovic, *Penn State U.*
May Berenbaum, *U. of Illinois*
Gabriele Bergers, *U. of California, San Francisco*
Bradley Bernstein, *Massachusettes General Hospital*
Peer Bork, *EMBL*
Bernard Bourdon, *Ecole Normale Supérieure de Lyon*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California Los Angeles* **(S)**
Christian Büchel, *U. Hamburg-Eppendorf*
Joseph A. Burns, *Cornell U.*
Gyorgy Buzsaki, *New York U. School of Medicine*
Blanche Capel, *Duke U.*
Mats Carlsson, *U. of Oslo*
David Clapham, *Children's Hospital Boston*
David Clary, *U. of Oxford*
Joel Cohen, *Rockefeller U., Columbia U.*
James Collins, *Boston U.*
Robert Cook-Deegan, *Duke U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Robert H. Crabtree, *Yale U.*
Roberta Croce, *Vrije Universiteit*
Janet Currie, *Princeton U.*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Frans de Waal, *Emory U.*
Stanislas Dehaene, *Collège de France*
Robert Desimone, *MIT*
Claude Desplan, *New York U.*
Ap Dijksterhuis, *Radboud U. of Nijmegen*
Dennis Discher, *U. of Pennsylvania*
Gerald W. Dorn II, *Washington U. School of Medicine*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
Christopher Dye, *WHO*
Todd Ehlers, *U. of Tuebingen*
David Ehrhardt, *Carnegie Inst. of Washington*
Tim Elston, *U. of North Carolina at Chapel Hill*
Gerhard Ertl, *Fritz-Haber-Institut, Berlin*
Barry Everitt, *U. of Cambridge*
Ernst Fehr, *U. of Zurich*
Anne C. Ferguson-Smith, *U. of Cambridge*
Michael Feuer, *The George Washington U.*
Toren Finkel, *NHLBI, NIH*
Kate Fitzgerald, *U. of Massachusetts*
Peter Fratzl, *Max-Planck Inst.*
Elaine Fuchs, *Rockefeller U.*
Daniel Geschwind, *UCLA*
Andrew Gewirth, *U. of Illinois*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *Rice U.*
Julia R. Greer, *Caltech*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zurich*
Kip Guy, *St. Jude's Children's Research Hospital*
Taekjip Ha, *U. of Illinois at Urbana-Champaign*
Christian Haass, *Ludwig Maximilians U.*
Steven Hahn, *Fred Hutchinson Cancer Research Center*
Michael Hasselmo, *Boston U.*
Martin Heimann, *Max-Planck Inst. Jena*
Yka Helariutta, *U. of Cambridge*
James A. Hendler, *Rensselaer Polytechnic Inst.*
Janet G. Hering, *Swiss Fed. Inst. of Aquatic Science & Technology*
Kai-Uwe Hinrichs, *U. of Bremen*
Kei Hirose, *Tokyo Inst. of Technology*
David Hodell, *U. of Cambridge*
David Holden, *Imperial College*
Lora Hooper, *UT Southwestern Medical Ctr. at Dallas*
Raymond Huey, *U. of Washington*
Steven Jacobsen, *U. of California, Los Angeles*
Kai Johnsson, *EPFL Lausanne*
Peter Jonas, *Inst. of Science & Technology (IST) Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kahne, *Harvard U.*
Daniel Kammen, *U. of California, Berkeley*
Masashi Kawasaki, *U. of Tokyo*
V. Narry Kim, *Seoul National U.*
Joel Kingsolver, *U. of North Carolina at Chapel Hill*
Robert Kingston, *Harvard Medical School*
Etienne Koechlin, *Ecole Normale Suprieure*
Alexander Kolodkin, *Johns Hopkins U.*
Alberto R. Kornblihtt, *U. of Buenos Aires*
Leonid Kruglyak, *UCLA*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Pennsylvania*
David Lazer, *Harvard U.*
Thomas Lecuit, *IBDM*
Virginia Lee, *U. of Pennsylvania*
Stanley Lemon, *U. of North Carolina at Chapel Hill*
Ottoline Leyser, *Cambridge U.*
Marcia C. Linn, *U. of California, Berkeley*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzan, *CIC biomaGUNE*
Jonathan Losos, *Harvard U.*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Laura Machesky, *CRUK Beatson Inst. for Cancer Research*
Anne Magurran, *U. of St. Andrews*
Oscar Marin, *CSIC & U. Miguel Hernández*
Charles Marshall, *U. of California, Berkeley*
C. Robertson McClung, *Dartmouth College*
Graham Medley, *U. of Warwick*
Tom Misteli, *NCI*
Yasushi Miyashita, *U. of Tokyo*
Mary Ann Moran, *U. of Georgia*
Richard Morris, *U. of Edinburgh*
Alison Motsinger-Reif, *NC State U.* **(S)**
Sean Munro, *MRC Lab. of Molecular Biology*
Thomas Murray, *The Hastings Center*
James Nelson, *Stanford U. School of Med.*
Daniel Neumark, *U. of California, Berkeley*
Kitty Nijmeijer, *U. of Twente*
Pär Nordlund, *Karolinska Inst.*
Helga Nowotny, *European Research Advisory Board*
Ben Olken, *MIT*
Joe Orenstein, *U. of California Berkeley & Lawrence Berkeley National Lab*
Harry Orr, *U. of Minnesota*
Andrew Oswald, *U. of Warwick*
Steve Palumbi, *Stanford U.*
Jane Parker, *Max-Planck Inst. of Plant Breeding Research*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Donald R. Paul, *U. of Texas, Austin*
John H. J. Petrini, *Memorial Sloan-Kettering Cancer Center*
Joshua Plotkin, *U. of Pennsylvania*
Albert Polman, *FOM Institute AMOLF*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
David Randall, *Colorado State U.*
Colin Renfrew, *U. of Cambridge*
Felix Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Jim Roberts, *Fred Hutchinson Cancer Research Ctr.*
Barbara A. Romanowicz, *U. of California,Berkeley*
Jens Rostrup-Nielsen, *Haldor Topsoe*
Mike Ryan, *U. of Texas, Austin*
Mitinori Saitou, *Kyoto U.*
Shimon Sakaguchi, *Kyoto U.*
Miquel Salmeron, *Lawrence Berkeley National Lab*
Jürgen Sandkühler, *Medical U. of Vienna*
Alexander Schier, *Harvard U.*
Randy Seeley, *U. of Cincinnati*
Vladimir Shalaev, *Purdue U.*
Robert Siliciano, *Johns Hopkins School of Medicine*
Denis Simon, *Arizona State U.*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
John Speakman, *U. of Aberdeen*
Allan C. Spradling, *Carnegie Institution of Washington*
Jonathan Sprent, *Garvan Inst. of Medical Research*
Eric Steig, *U. of Washington*
Paula Stephan, *Georgia State U. and National Bureau of Economic Research*
Molly Stevens, *Imperial College London*
V. S. Subrahmanian, *U. of Maryland*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Cambridge U.*
John Thomas, *North Carolina State U.*
Shubha Tole, *Tata Institute of Fundamental Research*
Christopher Tyler-Smith, *The Wellcome Trust Sanger Inst.*
Herbert Virgin, *Washington U.*
Bert Vogelstein, *Johns Hopkins U.*
Cynthia Volkert, *U. of Göttingen*
Douglas Wallace, *Dalhousie U.*
David Wallach, *Weizmann Inst. of Science*
Ian Walmsley, *U. of Oxford*
Jane-Ling Wang, *U. of California, Davis*
David A. Wardle, *Swedish U. of Agric. Sciences*
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Ian A. Wilson, *The Scripps Res. Inst.* **(S)**
Timothy D. Wilson, *U. of Virginia*
Rosemary Wyse, *Johns Hopkins U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Pennsylvania School of Medicine*
Jonathan Zehr, *U. of California, Santa Cruz*
Len Zon, *Children's Hospital Boston*
Maria Zuber, *MIT*

BOOK REVIEW BOARD

David Bloom, *Harvard U.*, **Samuel Bowring,** *MIT,* **Angela Creager,** *Princeton U.*, **Richard Shweder,** *U. of Chicago,* **Ed Wasserman,** *DuPont*

Published by AAAS

EDITORIAL

Science, all inclusive

After years of public leadership around fairness, equity, and diversity in science and engineering, the American Association for the Advancement of Science (AAAS) and its journal *Science* recently have been questioned about this leadership. AAAS is, after all, the organization that stopped all annual meetings in the "Jim Crow" South in the 1950s; that, at great expense, pulled its meeting out of Chicago to respect the ban on meeting in states that did not ratify the Equal Rights Amendment (1979); that held the first fully accessible professional meeting in Boston (1976), long before passage of the Americans with Disabilities Act; and that, in 1972, established the Committee on Opportunities in Science to advise what is now a mostly female AAAS Board of Directors on diversity issues. Long before public support of the rights of LGBTQ individuals, AAAS passed a nondiscrimination resolution (1975). The values upon which this organization is based not only remain intact, but continue to strengthen through dialogues with the scientific community and through actions that advance science for the benefit of all.

"How do we turn workplaces...into places where...civility and freedom from harassment are the norms?"

Last month, a letter to AAAS and *Science* leadership* noted recent incidents in which content in *Science* and *Science* Careers reinforced damaging stereotypes about underrepresented groups in science, and suggested greater scrutiny of published and posted material. Indeed, as *Science*'s Editor-in-Chief Marcia McNutt pointed out,† the journal regrets these unfortunate incidents and is pursuing better oversight strategies, including diversity sensitivity training for staff. While *Science* looks internally to make improvements, AAAS continues to look outward to its society colleagues to discuss larger structural barriers to quality and diversity in science.

Those, like me, who work directly on these barrier issues, have focused on keeping track of the numbers, and use increases in the participation of women and minorities as indicators of reduced obstacles. There has been an expectation that a "critical mass" might promote a welcoming climate and culture. Although this is happening in some scientific fields and institutions, it is not the norm. Other advocates for inclusion are looking at the science itself, questioning whether producing the best science should include an awareness of diversity when considering the way that work is done, the scientific questions asked, or the research agendas set.

The science community prizes objectivity, but research indicates that this isn't necessarily reflected in the behavior and choices of scientists. Stereotypes and implicit bias can infect everyone, causing people to first be seen as minorities or women rather than as scholars, peers, and colleagues. Disability or gender identity is recognized before ability. How do we turn workplaces, institutions, and departments into places where expectations of civility and freedom from harassment are the norms?

Yes, leadership matters, and leaders need to find out how different groups experience their departments and institutions. Leaders must look at the policies, structures, and behaviors in workplaces, not just learned responses to these. Are people being forced into "no-win" choices? A two-career couple who elects to put his career advancement above hers may be responding to the sad reality that he will likely be offered a higher salary than she will. Unfortunately, for them, this may be a rational decision—not a fair one, but understandable. Who is watching to ensure that negotiations for salary and research start-up packages don't leave women with less than they deserve? Gender and minority awareness and fairness must be treated in the same way as research integrity, including discussion of bystander responses to unfairness and abuse.

The larger science community must examine all its institutions and workplaces in light of today's changing cultural, educational, and business landscape, accepting responsibility to call out unfairness whenever and wherever it appears. We can only address collective global challenges if we disconnect from the structures of the past that are hobbling the ability to move forward together.

– Shirley M. Malcom

Shirley M. Malcom is the director of Education and Human Resources Programs at AAAS, Washington, DC; co-chair of the Gender Advisory Board of the United Nations Commission on Science and Technology for Development; and co-chair of Gender InSITE, a global initiative. E-mail: smalcom@aaas.org

*www.buzzfeed.com/catferguson/science-journal-reinforces-dangerous-stereotypes
†http://scim.ag/EN0716; http://scim.ag/WLC0710

10.1126/science.aad1421

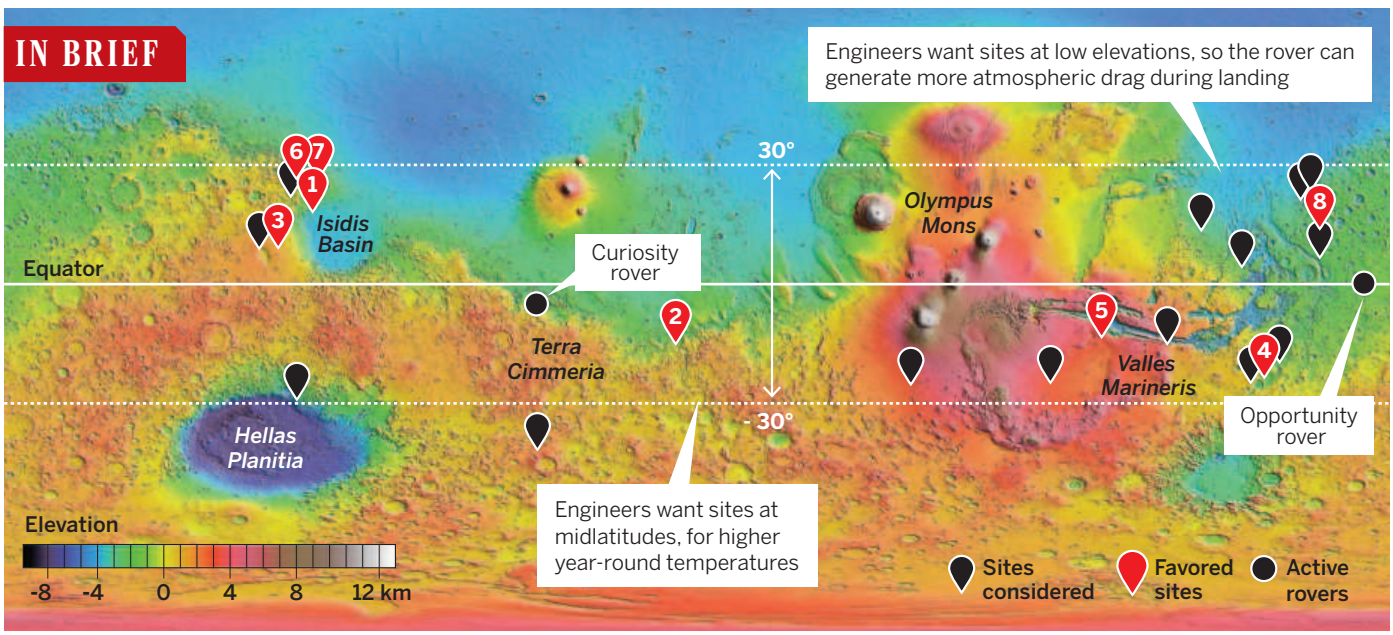
IMAGES: (TOP RIGHT) AAAS; (INSET) ADAPTED FROM LEONTURA/ISTOCK AND KUBKOO/ISTOCK

Published by AAAS

NEWS

> “It’s a new and emerging subdiscipline ... a sign of the times.”
>
> **Archaeologist Martin Callanan** of the Norwegian University of Science and Technology in Trondheim on “glacial archaeology,” a growing field as mountain climbers’ remains, once buried in glaciers, are increasingly revealed by melting ice.

IN BRIEF



Top eight sites (by average vote): **1.** Jezero Crater | **2.** Columbia Hills | **3.** NE Syrtis | **4.** Eberswalde | **5.** SW Melas | **6.** Nili Fossae Trough | **7.** Nili Fossae Carbonate | **8.** Mawrth

Ancient river deltas, hot springs beckon next Mars rover

Researchers have narrowed the list of possible destinations for NASA’s next Mars rover, due for launch in 2020. Last week they voted for eight sites with high scientific potential. Jezero Crater, a site that holds the remains of a river delta, was the top vote getter; Columbia Hills, a site previously explored by the defunct rover Spirit, came in second, because of silica deposits that indicate an ancient system of hot springs. The $1.5 billion rover is supposed to gather more than 30 rock samples and lay them on the ground in metal tubes; subsequent missions will pick them up and return them to Earth. Scientists want to collect igneous samples, which can help explain how Mars formed, as well as rocks that have the potential for preserving signs of life. The 2020 rover will land using a “sky crane” system similar to that of the Curiosity rover, and engineers will evaluate the relative hazards before a final site is chosen. **http://scim.ag/marstargets**

GRAPHIC: ADAPTED FROM NASA/MOLA SCIENCE TEAM BY A. CUADRA/*SCIENCE*

AROUND THE WORLD

Neutrino experiment’s first data

BATAVIA, ILLINOIS | The $278 million NOvA experiment at Fermi National Accelerator Laboratory here has collected only 8% of its expected data, but early results suggest it is on course to achieve its main goal of ranking mysterious particles called neutrinos by their weights. Nearly massless and barely interacting with other matter, neutrinos come in three “flavors”—electron, muon, and tau—that morph into one another. Physicists don’t know whether there are two light ones and one heavy one (the so-called normal hierarchy) or the other way around (inverted hierarchy). NOvA aims to find out by firing muon neutrinos 810 kilometers through Earth to a 14,000-tonne detector in Minnesota and spotting electron neutrinos emerging in the beam. The first year of data, yielding between six and 11 conversions, hints that the hierarchy is normal, NOvA physicists reported last week. Another key type of experiment to test whether the neutrino is its own antiparticle may be feasible only if the hierarchy is inverted.

Concerns over larger Suez Canal

SUEZ CANAL, EGYPT | An $8.5 billion expansion of the Suez Canal adds 35 kilometers of new channels, expanding ship capacity by 12% and decreasing wait time

Published by AAAS

Scientists worry the Suez Canal's expansion will increase invasive species.

by 8 hours. But scientists say the expansion, which Egypt celebrated last week, is also playing "ecological Russian roulette" with the Mediterranean Sea because of the canal's increased likelihood of serving as a conduit for invasive species from the Red Sea and Indian Ocean. The Egyptian government has yet to carry out a thorough environmental risk assessment, they complain. Scientists would also like to see an effective monitoring, early detection, and rapid response system put in place. http://scim.ag/suezexpansion

Alzheimer's data dispute heats up

LOS ANGELES, CALIFORNIA | Each side in a bitter legal dispute between the University of California, San Diego (UCSD), and the University of Southern California (USC) over control of data from a large, federally funded Alzheimer's study scored a victory last week. On 4 August, a California judge ordered USC, the new home of researcher Paul Aisen, to restore control of all Alzheimer's Disease Cooperative Study (ADCS) electronic data and computer systems to UCSD, where Aisen had worked. On the same day, pharmaceutical company Eli Lilly announced that it is ending its contract with UCSD for one clinical drug trial within the network and moving it to USC. The court injunction can't stop the company from moving its own data to USC, says Russ Barton, chief operating officer for the Alzheimer's platform team at Lilly. But it could prevent computer systems developed at UCSD from being transferred, according to UCSD lawyer Dan Sharpe, who says the university wants to recover its costs.

NEWSMAKERS

HHMI chief departing in 2016

Biochemist **Robert Tjian**, president of the Howard Hughes Medical Institute (HHMI), announced that he will step down at the end of 2016 after nearly 8 years as president of the $18.6 billion research foundation in Chevy Chase, Maryland. The largest private funder of U.S. biomedical research, HHMI supports more than 330 investigators based at major U.S. universities. Tjian shored up HHMI support for early career scientists, created a 5-year award to help senior investigators phase out of the program, revamped college education programs, and launched a science documentary company. He also oversaw new collaborations with other foundations, including a plant science program and the start-up of *eLife*, an open access journal. Tjian, who turns 66 next month, plans to return to his HHMI-supported lab studying gene regulation at the University of California, Berkeley. http://scim.ag/TjianHHMI

Obama makes NSF, DOE picks

Last week President Barack Obama nominated **Richard Buckius** to fill the job he's informally held since arriving at the National Science Foundation (NSF) last summer, asking the U.S. Senate to make him the agency's deputy director. Buckius, a mechanical engineer, worked under NSF Director France Córdova at Purdue University and has held senior management positions at NSF. Obama also tapped physicist **Cherry Murray** of Harvard University to head the Department of Energy's (DOE's) $5 billion Office of Science. Murray's nomination follows the U.S. Senate's failure to act on Obama's 2013 pick for the DOE post, physicist Marc Kastner, then of the Massachusetts Institute of Technology. Murray spent her early career at Bell Laboratories and was dean of Harvard's engineering school from 2009 to 2014.

Who will be the next MacGyver?

In March, Hollywood bigwigs and engineering institutions dangled a tantalizing proposition before aspiring screenwriters: the chance to create a new TV series starring a female hero doing science. On 28 July, 12 finalists pitched their series ideas—part of a pilot script, along with concept art—to an audience of Hollywood producers, actors, and other insiders; five winners were chosen. Each will receive $5000 and a chance to work with Hollywood mentors to develop their series. The winning concepts ranged from spy action to high school comedy to historical. (Above is the concept art for *Ada and the Machine*, a fictionalized retelling of the story of Ada Lovelace, the daughter of poet Lord Byron, who in 1832 wrote logarithms and created programs for mathematician Charles Babbage's new "calculating machines.") We asked *Science* readers to vote on which of the five winners they prefer—and the results are in. The winner, with 33% of the vote, is an action series called *Q Branch*, about a woman who invents superspy gadgets. http://scim.ag/nextmacgyver

IMAGES: (TOP TO BOTTOM) ANDREJ67/ISTOCKPHOTO; ZOE CHEVAT

Published by AAAS

SCIENCE DIPLOMACY

Iran's atomic czar describes the art of the deal

Ali Akbar Salehi tells *Science* how he helped seal the Iran nuclear agreement

By **Richard Stone**, *in Tehran*

Last February, nuclear talks between Iran and world powers were foundering. The two sides had found common ground on the deal's broad outlines, but the devil lay in the technical details. Stymied, Iranian officials asked their top nuclear scientist to join the talks: Ali Akbar Salehi, president of the Atomic Energy Organization of Iran (AEOI).

In an exclusive interview with *Science* at AEOI headquarters in north Tehran, Salehi, 66, related how he would only agree if his opposite number in the United States, Energy Secretary Ernest Moniz, sat across from him at the table. The result was several weeks of intense science diplomacy between the two physicists, who overlapped at the Massachusetts Institute of Technology (MIT) in Cambridge in the mid-1970s, when Salehi was earning a Ph.D. there in nuclear engineering. They helped overcome technical obstacles, and last month Iran and the P5+1—the United States and its five allies—reached an agreement that is designed to block Iran's paths to a nuclear weapon in exchange for a gradual lifting of sanctions imposed as a result of Iran's nuclear program.

As the U.S. Congress reviews the deal, the battle for public opinion is heating up. In a letter to President Barack Obama on 8 August, 29 prominent U.S. physicists and nonproliferation experts (including *Science*'s executive publisher, Rush Holt), praised the "innovative agreement" for providing "much more stringent constraints than any previously negotiated non-proliferation framework." In a 7 August interview with *Der Spiegel*, Israel Defense Minister Moshe Yaalon called the deal "a historic mistake" that will "allow Iran to become a military nuclear threshold state."

Framed photos of five Iranian nuclear scientists assassinated over the past decade hang in Salehi's office, with a bouquet of red roses set on a table below. Salehi claims their deaths emboldened, rather than deterred, Iran's nuclear establishment, and insists that his country's enrichment program is intended only to produce fuel for civilian power reactors. This transcript was edited for brevity and clarity; the full interview is available online. http://scim.ag/Salehi

Q: What stands out as the most memorable moment you spent with Secretary Moniz during the negotiations?

A: One came when I asked Moniz in our last days in Vienna, "Do you have any news about Professor Mujid Kazimi?" He's a professor of nuclear engineering at MIT. The next day, Moniz comes to me and says, "I have bad news for you." I thought it's about the negotiations. I said, "What is it?" He said, "Mujid has passed away." I said, "My God. Yesterday we were talking about him and you said he's fine." I requested Moniz to see if we can get his family's telephone number. And he found it and gave it to me, and I called the family. I expressed my condolences.

We know each other through all these [mutual] friends. It was very natural for him to look out for his country's interests. But this did not prevent either of us from being rational. We tried to be logical and fair. We understood each other. I understood his constraints. He understood mine. That's how we could move forward.

Q: How did you and Moniz end up as negotiating partners?

A: The negotiations were stalled. It was about a year and a half that they were negotiating, and they couldn't move forward. In February, I was summoned by my superiors. One of the officials said, "Well, Ali, you have to join the negotiations." I said, "You're joking. What can I do? [Foreign Affairs Minister Mohammad Javad] Zarif has

PHOTO: AP PHOTO/KEYSTONE, SALVATORE DI NOLFI

Published by AAAS

done his best. What more can I do?"

But my superiors insisted. I said, "OK, I will go, but on condition that my counterpart also participates." They said, "Who is your counterpart?" I said, "The secretary of the DOE [Department of Energy]." Our side contacted [chief U.S. negotiator] Wendy Sherman, saying "Look, we are intending to bring Mr. Salehi with us. But on condition that the DOE secretary also joins." After a few hours, Sherman responded and said, "We welcome this and we'll bring along Secretary Moniz." [laughs] And when I heard the news that he's coming, I said, "OK, I will go." [A DOE spokesperson confirmed Salehi's account.]

I thought I was going on a mission impossible. I'm so happy that the final outcome made all of us happy. Yes, we have some constraints, but what are those constraints? From the American perspective, they are to prevent us from diverting to nonpeaceful activities. But then we never ever had this idea of diverting to nonpeaceful activities.

Q: You're said to have a close relationship with the supreme leader, Seyyed Ali Khamenei.
A: Well, I wouldn't say we have a special relationship, but the supreme leader knows me because I have been in different government responsibilities since the revolution started. I was chancellor of a university. I was deputy minister of higher education two times. I'm happy that the supreme leader, yes, has put his trust in me.

Q: One of the most contentious issues in the negotiations was R&D on advanced centrifuges. From your perspective, what was hard to accept?
A: When you are negotiating, each party is trying its best to have the bigger piece of the cake. That's very natural. The Americans said, "If there's no research on centrifuges, we will be very happy." We said, "We would not be happy. We understand you have some concerns. Let's see how we can mitigate them." Neither side got the ideal it was looking for. We met in the middle.

Q: Your program is slowed.
A: Yeah. If we were free, probably within 8 years we could have come up with a cascade of 164 [advanced IR8 centrifuges]. That is not a big constraint for us, but that pleases the other side so we said, "OK." It could have been better, of course, if we had a bigger cascade because then even the process of enrichment would have been assessed, not only the mechanical characteristics of the machines.

We do not take that as a constraint. So I would say on R&D, the apparent limitations that we have accepted, that we have agreed to, it's not really a limitation.

Q: So there's nothing in particular in the sphere of R&D that you consider a huge sacrifice for the sake of the pact?
A: I don't think so. We would be working on different advanced machines. We would be working on the IR8, on the IR6. The IR8 and IR6 are the two candidates that could really meet our needs in terms of producing enough enrichment capacity to meet the annual needs of [the Bushehr power reactor]. And 10 years from now, we will have two other nuclear power reactors. But using [the permitted] centrifuges, in 15 years we will be in a position to meet the fuel requirements of these reactors.

Salehi's sparring partner: U.S. Energy Secretary Ernest Moniz.

PHOTO: AP PHOTO/ANDREW HARNIK

Q: U.S. negotiators were concerned about other forms of enrichment, like laser enrichment. Was Iran pursuing these?
A: No. We did some laser enrichment in the past. We informed the IAEA [International Atomic Energy Agency] and dismantled the equipment.

Q: Siegfried Hecker, the former director of Los Alamos National Laboratory, told me he was surprised Iran would agree to a blanket prohibition on studying metallic uranium and plutonium, because there are other uses for these, not just for weapons.
A: Yes. We had a lengthy discussion on that in the negotiations. Depleted uranium metal can be used as shielding, for example. I insisted that we would have to make an exception for depleted uranium. Then the other side said, "Look, we need to tell our officials that we have blocked all the pathways to weapons production. And one of the pathways is metal production."

I finally accepted that. We do not intend to enrich uranium to 90%, so we will not have 90% enriched uranium to turn into metal. We said that if we need some depleted uranium, we may ask for it from [the P5+1] during this period. But we will not produce the metal. We did [produce] plutonium. And we told the IAEA, that's it. That was in 2003.

Q: As part of the agreement, the Fordow enrichment facility will be turned into an international research center. What do you have in mind?
A: It's very difficult to say offhand. I would have to discuss it with my colleagues and with colleagues from the P5+1.

Q: You've already hosted Russian scientists. They're going to help modify the uranium centrifuges to produce stable isotopes for industry.
A: That is for sure. Fordow has two wings. Part of one wing will be dedicated to stable isotopes. That is already agreed upon.

Q: Fordow is a military site. Is the military going to easily relinquish it?
A: It's not controlled by the military. Decades ago, it was a place where the military stored ammunition.

Q: The nuclear agreement calls for increased Iranian participation in ITER, the international fusion experiment. Does Iran have a fusion research program?
A: Yes. Near where you're sitting, we have three tokamaks. We are one of the leading countries in West Asia working on fusion. This is my second time heading the Atomic Energy Organization. In my previous appointment, I made fusion our essential goal. It was given our highest priority because fusion is the future source of energy.

Q: AEOI went through some very dark days a few years ago, when five nuclear scientists were assassinated. Do their deaths cast a shadow on international collaboration?
A: No. We have a very peculiar characteristic of our nation. Being Muslims, we are ready for any kind of destiny because we do not look upon it like you have lost your life. OK, but you have gained martyrdom and we believe in eternity.

For our people, it's easy to absorb such things. I mean, this did not really turn into an impediment to our nuclear activities. In fact it gave an impetus to the field, in the sense that after [the assassinations], many students who were studying in other fields changed to nuclear science.

Q: What do you want to be remembered for?
A: As a person who did good for mankind. That's it. ■

Published by AAAS

Genomics of cold-adapted creatures such as Antarctic icefish is a major theme of the new report.

POLAR SCIENCE

Report sets new goals for U.S. Antarctic Program

Committee pushes study of ice loss, genomics, and radiation from the beginning of the universe

By **Carolyn Gramling**

The United States' research in Antarctica needs fresh initiatives and better equipment, a new report by a committee of the National Academies concludes. But how to afford them remains a conundrum.

The report—commissioned by the National Science Foundation (NSF), which funds the United States Antarctic Program (USAP)—is the third major assessment of the program in 5 years, aiming to streamline the program in an era of relatively flat budgets and rising infrastructure costs. It builds on a 2011 report by the National Research Council, which identified important areas of future research for Antarctica and the Southern Ocean. This time, NSF asked the committee to lay out a strategic vision for research on the continent over the next decade, identifying specific research priorities while taking into account the program's logistical needs. NSF and its Division of Polar Programs invest about $70 million a year in science and about $255 million in infrastructure and logistics.

"Earlier reports were blue-sky," supporting curiosity-driven research on the continent without setting priorities, says Robin Bell, a co-chair of the report and a glaciologist at Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York. The new report, which takes into account input from more than 450 scientists in the Antarctic research community, does call for continuing a "core program" of such investigator-driven research. But it also advocates for the creation of three priority research initiatives.

The Changing Ice Initiative would fund research addressing both what's driving Antarctica's ice mass loss and its future course, as well as how it could contribute to global sea-level rise. The Antarctic Genomics Initiative would take stock of the uniquely isolated and adapted life in the region. "It's a perfect lab for looking at organisms in extreme conditions, and for how they'll change in a changing environment," Bell says. The third priority area would create a new, next-generation Cosmic Microwave Background (CMB) program to probe deeper into the origins of the universe. A follow-on to the research already conducted at the BICEP telescope at the South Pole, the proposed program would include a family of telescopes at the South Pole and Chile.

In choosing these priorities, Bell says, the committee looked for topics with compelling science, high potential for societal impact, and high "partnership potential" within NSF, between NSF and other agencies, and with international partners. Projects that would require too many resources were rejected. In particular, a proposal to build a second-generation IceCube neutrino observatory did not make the list because of concerns that it would starve other projects of logistical support.

The priorities closely overlap with six overarching research priorities outlined last year in *Nature* by the Scientific Committee on Antarctic Research (SCAR), a group of scientists and policymakers from 22 countries. "Overall there's a great confluence of thinking, within the U.S. and internationally, about Antarctica and Southern Ocean science priorities," says Chuck Kennicutt, an oceanographer at Texas A&M University in College Station, who is also a former president of SCAR.

But Kennicutt notes that it's not clear how these projects would be funded within NSF's constrained Antarctic budget. And parceling out funds between the new theme-driven proposals and principal investigator–driven projects will be only one part of NSF's looming Antarctic challenge. Conducting research in Antarctica is already expensive—and infrastructure upgrades are sorely needed. In 2012, a blue-ribbon panel convened by NSF identified multiple areas of Antarctic infrastructure requiring improvement (*Science*, 27 July 2012, p. 397).

The new report adds to this sense of urgency, calling for better weather forecasting, improved overland and air access to remote field sites in the deep interior of the continent, and greater information technology capabilities. Data transmission capacity from the South Pole station is already inadequate, and the proposed next-generation CMB program will increase the station's bandwidth needs further. And then there's ship support: The United States has only one heavy icebreaker, the 40-year-old *Polar Star*, capable of clearing thick ice from McMurdo Sound to deliver supplies to the station there. The U.S. program has had to rely on foreign icebreakers for this in some past seasons.

But the priority setting should help researchers make the case for more investment. Last year's SCAR report and the new National Academies report, Kennicutt says, are forcing researchers to "decide what's important" and to justify their work to a broader audience. "These reports lay out why this is compelling science, why we want to spend the money it costs to be in Antarctica," he adds. "They show that the community is organized." ■

PHOTO: UWE KILS/WIKIMEDIA COMMONS

Published by AAAS

SYNTHETIC BIOLOGY

Modified yeast produce opiates from sugar

Engineered microbes raise hopes for better medicines and concerns about illicit drugs

By **Robert F. Service**

Move over, poppies. In one of the most elaborate feats of synthetic biology to date, a research team has engineered yeast with a medley of plant, bacterial, and rodent genes to turn sugar into thebaine, the key opiate precursor to morphine and other powerful painkilling drugs that have been harvested for thousands of years from poppy plants. The team also showed that with further tweaks, the yeast could make hydrocodone, a widely used painkiller that is now made chemically from thebaine.

"This is a major milestone," says Jens Nielsen, a synthetic biologist at Chalmers University of Technology in Göteborg, Sweden. The work, he adds, demonstrates synthetic biology's increasing sophistication at transferring complex metabolic pathways into microbes.

By tweaking the yeast pathways, medicinal chemists may be able to produce more effective, less addictive versions of opiate painkillers. But some biopolicy experts worry that morphinemaking yeast strains could also allow illicit drugmakers to brew heroin as easily as beer enthusiasts home brew today—the drug is a simple chemical conversion from morphine. That concern is one reason the research team, led by Christina Smolke, a synthetic biologist at Stanford University in Palo Alto, California, stopped short of making a yeast strain with the complete morphine pathway; medicinal drug makers also primarily use thebaine to make new compounds.

Synthetic biologists had previously engineered yeast to produce artemisinin, an antimalarial compound, but that required inserting just a handful of plant genes. To get yeast to make thebaine, Smolke's team coaxed the cells to express 21 genes in total, including many added from a diverse set of species (see graphic); making hydrocodone took 23 genes.

Their success, reported online this week in *Science*, caps a race to install the complex opioid pathway in yeast. Last year, Smolke's team reported engineering yeast to carry out the tail end of the process, going from thebaine to morphine. In April, Vincent Martin, a microbiologist at Concordia University in Montreal, Canada, and his colleagues said they had created yeast that could go from an earlier intermediate compound called R-reticuline to morphine. A few weeks later, John Dueber, a synthetic biologist at the University of California, Berkeley, and colleagues announced yeast that carries out most of the first half of the pathway, going from glucose to another intermediate compound, S-reticuline. Finally, two groups reported in late June that they had identified the long-sought enzyme needed to carry out the chemical transformation in the middle, S-reticuline to R-reticuline.

Even so, many predicted it would take years to put all the pieces together. As it turns out, back in May, Smolke and her colleagues had already largely finished the task. "It shows this field is really moving fast," says Kenneth Oye, a biotechnology policy expert at the Massachusetts Institute of Technology in Cambridge.

The most important challenge, Smolke says, was increasing the efficiency of each step so losses wouldn't build up. In one step, for example, a plant enzyme called SalSyn was doing a poor job of converting R-reticuline to another compound called salutaridine. Eventually, Smolke's team discovered that the yeast made the enzyme incorrectly, attaching the wrong sugars to it. The researchers fixed the problem by reengineering the inserted plant gene.

Smolke plans to go on tinkering. The microbes need to increase output of thebaine by a factor of 100,000 for drug companies to be interested in using them to make medicines. That won't be easy. But Martin notes that researchers boosted the output of the artemisininmaking yeast by a similar amount. "It will happen," he says. "The only question is how fast." Smolke recently formed a company called Antheia, based in Palo Alto, that aims to push that pace.

To keep up with the yeast engineers, Oye says policy experts need to develop rules to limit the risk of unintended uses of engineered microbes. In the case of opiatemaking yeast, such rules might forbid developing strains to produce illicit drugs, such as heroin, and require scientists to build in genes that prevent the microbes from living outside of a controlled laboratory environment.

Not everyone is worried about home-brewed opiates. Andrew Ellington, a synthetic biologist at the University of Texas, Austin, calls such fears "overblown." The idea that producing vanishingly small quantities of opiates through fermentation is somehow going dwarf the problem of illegal drugs made from poppies is "laughable," he says. But Martin disagrees. "Poppy fields are not readily available to someone in Chicago, whereas yeast can be made available to anyone." ■

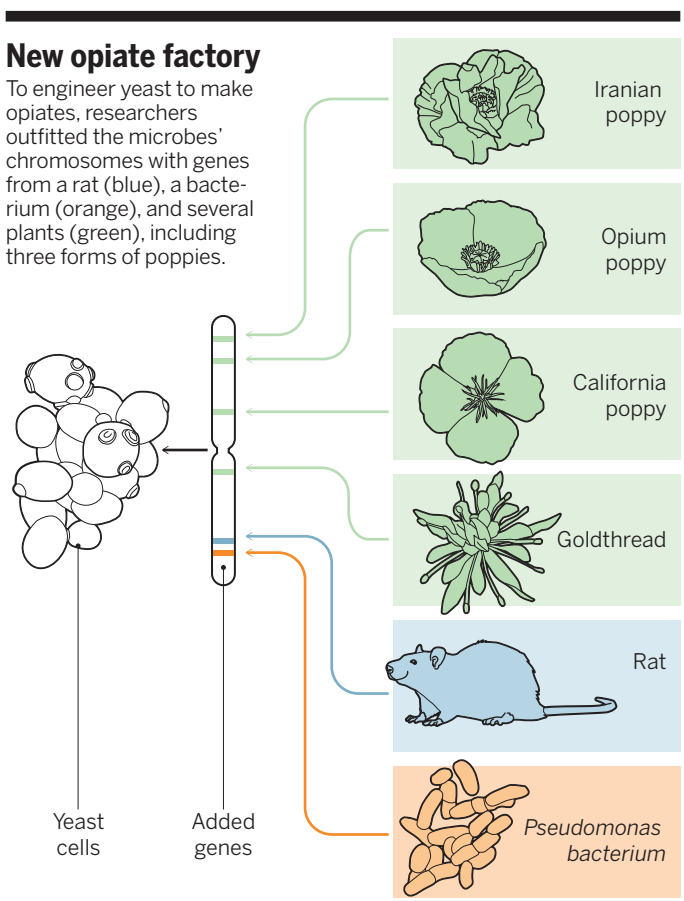


ILLUSTRATION: A. CUADRA/*SCIENCE*

Published by AAAS

MARINE GEOLOGY

Russian claim heats up battle to control Arctic sea floor

New data on undersea mountains support conflicting claims to the North Pole

By **Lucas Laursen**

A long-simmering struggle over who owns the Arctic sea floor intensified last week, as Russia submitted an updated territorial claim—together with new seafloor maps and samples to support it. Russia's claim to an additional 1.2 million square kilometers of seabed near the North Pole sets up a potential clash with other Arctic nations. Denmark has asserted ownership of part of the area claimed by Russia, and Canada is also expected to file an overlapping claim.

The competing submissions represent "a battle of the countries' ambitions" to control the Arctic, and an effort to capture "the North Pole brand," says geophysicist Nina Lebedeva-Ivanova of the University of Oslo. And they are sure to fuel technical debates, because the United Nations Convention on the Law of the Sea (UNCLOS), which entered into force in 1994, links territorial claims to the fine points of undersea geology.

The law gives each coastal nation control over an exclusive economic zone (EEZ) that extends 370 kilometers off its shoreline. But if a country can provide geological proof that the continental shelf that slopes seaward from its shore extends beyond the EEZ, it can expand its claim to seafloor resources, including oil, gas, and minerals (but not sea life in the water column), by up to an additional 278 kilometers. This "extended continental shelf" provision has enabled some nations, including Australia and New Zealand, to expand their seabed territories by up to 50%.

Some of the five Arctic nations—Russia, Canada, the United States, Denmark, and Norway—are angling for similar gains. (The United States has not ratified UNCLOS, and so cannot participate in the claims process.) In 2001, Russia staked a claim to a huge chunk of the polar sea floor, stretching to the North Pole, arguing that two underwater mountain ranges—the Lomonosov and Mendeleev ridges—are extensions of its continental landmass. But in 2002, the United Nations's Commission on the Limits of the Continental Shelf, a 21-member technical body that evaluates the science underlying such claims, recommended that Russia revise and resubmit (*Science*, 6 December 2002, p. 1877).

That decision helped accelerate efforts by Arctic nations to map and sample the little-studied polar sea floor, much to the delight of geoscientists (*Science*, 16 March 2007, p. 1525). Ships equipped with sonars and seismic mapping systems—which use sound to chart seafloor geology—crisscrossed the Arctic, aided at times by warmer summers that opened ice-locked waters. "Fifteen years ago I would have said [such a concerted effort] was impossible," says geophysicist Wilfried Jokat of the Alfred Wegener Institute in Bremerhaven, Germany.

Norway has already tapped such data to back a successful 2009 continental shelf claim. Other new data are behind the competing Danish and Russian claims. Last year, Denmark together with Greenland (which Denmark represents in foreign affairs) argued that the Lomonosov Ridge, which bisects the Arctic, is an extension of Greenland's continental shelf. They laid claim to 895,000 square kilometers of sea floor, extending past the North Pole into areas previously claimed by Russia and Canada (see map).

Russia's 3 August claim, by contrast, argues that both the Lomonosov Ridge and the roughly parallel Alpha-Mendeleev ridge system are extensions of Russia's continental shelf. And it says analyses of 120 rock samples, including basalt taken from the Mendeleev Ridge, will help back that claim.

For researchers, the submissions are tantalizing. They include new data that could help resolve long-standing questions, such as whether the Alpha-Mendeleev ridge system is part of a continent or the oceanic crust, Lebedeva-Ivanova says. She's especially interested in seeing Russia's basalt data. But there's the rub: The United Nations allows the full submissions to remain secret. "Nobody's seen that data," so it is hard for outside experts to weigh in, says marine geologist and consultant Lindsay Parson in Romsey, U.K.

Still, researchers have some informed opinions. Germany's Jokat has collected data that support the idea that the Mendeleev Ridge is part of the ocean crust. If the U.N. experts agree, that could weaken Russia's case for a big chunk of its claim. In contrast, there's consensus that the Lomonosov Ridge is of continental origin—potentially giving support to both Russia and Denmark. Whatever the U.N. panel decides, however, its views are advisory only; it's up to diplomats to set final marine boundaries.

It could take years for the experts to issue their opinions. In part, Lebedeva-Ivanova says, that's because UNCLOS includes "strange," outdated, and possibly misleading geological terminology. In the meantime, scientists fear funding to study the Arctic sea floor will decline as the competition to claim the North Pole winds down. ■

Lucas Laursen is a freelance journalist in Madrid.

Border dispute

Both Russia and Denmark say the Lomonosov Ridge is connected to their continental shelf, leading to overlapping seabed claims.

DATA: GEOLOGICAL SURVEY OF DENMARK AND GREENLAND (GEUS)

Published by AAAS

Members of Peru's isolated Mashco Piro tribe have frequently appeared on a riverbank and encountered outsiders.

PHOTO: © JEAN-PAUL VAN BELLE

ANTHROPOLOGY

Mashco Piro tribe emerges from isolation in Peru

Crisis sparks debate about whether to make official contact or leave isolated tribespeople alone

By **Andrew Lawler**

The Peruvian government announced last month that it would attempt to help and potentially contact an isolated indigenous tribe that lives deep in the Amazon. The tribe has been sighted more and more frequently, and its members have killed two local villagers. But some fear the plan could further jeopardize the group, and on 3 August the government backtracked on the implications of its announcement, insisting that it would not make the first move toward official contact.

The case highlights an ongoing international debate about how to best help emerging tribes, who lack immunity to common diseases and are among the most vulnerable people on the planet (*Science,* 5 June, p. 1072). "We are extremely worried about this situation and its possible disastrous consequences," says Francisco Estremadoyro, director of Lima-based ProPurús, a nonprofit organization that seeks to protect isolated peoples and the environment in eastern Peru.

He and others are concerned about 30 members of the Mashco Piro tribe in Manú National Park, a remote forested area bordered by the Madre de Dios River. Peruvian policy is to avoid all contact with such isolated tribes and protect them from intruders on their reserves. But in practice this policy is difficult to carry out.

Anthropologists report more than 100 sightings of these tribespeople since 2014. The Mashco Piro raided a nearby village for machetes and goods and killed two villagers, the most recent in May. In past months they have also repeatedly gestured at, called to, and received goods from local people. Tour operators sell tickets for "human safaris" along the river, and missionaries are reported to have given food and clothing to the group in the past year. "You can see a group on the beaches for hours, waiting for the boats to pass," writes Luis Felipe Torres, a Peruvian ministry anthropologist who has spent time in the area, in an essay published 22 July. "They are especially interested in bananas, cassava, sugarcane, machetes, and pots." Torres concludes that "they are deliberately seeking to interact with people transiting the river."

Given such credible reports, "there are no reasonable grounds to interpret this behavior as a sign that this group wants to remain unconnected to the rest of society," said Patricia Palacios Balbuena, vice minister in Lima's Ministry of Culture that oversees tribal affairs, in a 21 July statement. The ministry subsequently approved a 6-month plan to increase patrols, discourage raids, and make contact with the Mashco Piro "only if they make an appearance and show a willingness for a conversation." But a ministry official also told reporters that the government planned "controlled contact," raising concerns among some nonprofit groups.

When and how to make contact with isolated tribes is a hotly contested issue. Anthropologists Kim Hill of Arizona State University, Phoenix, and Robert Walker of the University of Missouri, Columbia, recently proposed in an editorial (*Science*, 5 June, p. 1061) that "a well-designed contact can be quite safe," but the nonprofit group Survival International in London accused them of promulgating a "dangerous and misleading" idea. Without medical care, isolated tribes can be all but wiped out by the flu or other diseases.

In the Mashco Piro's case, given the contact already occurring, infection is likely, officials say. In the initial announcement, Balbuena called for "immediate action by the competent authorities to safeguard their health and prevent negative consequences of uncontrolled contact."

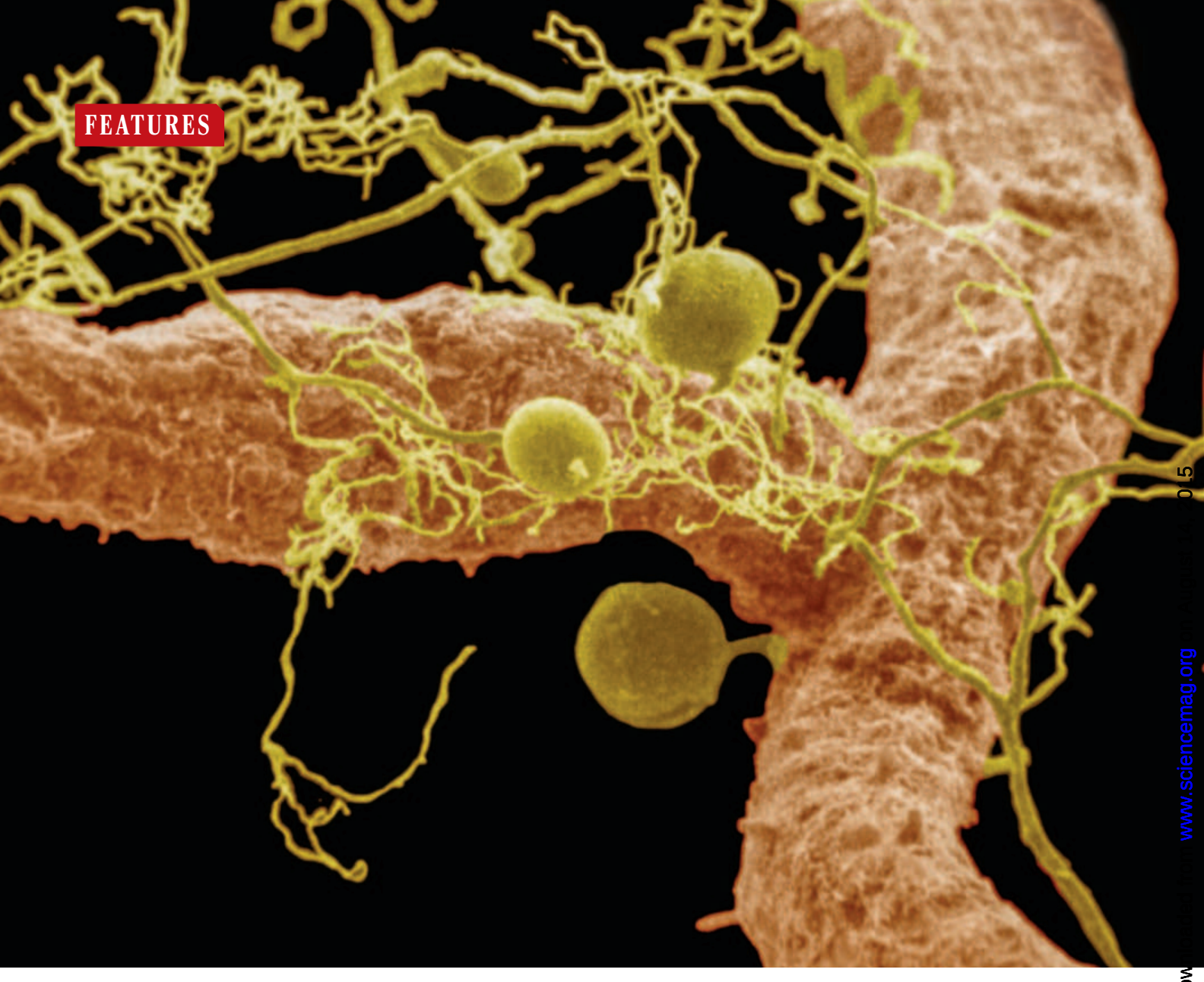
But Survival International and ProPurús protested Balbuena's statement and the vague plan, fearing that it could set a dangerous precedent. Rebecca Spooner, campaigns officer at Survival International, insists that the Mashco Piro's conflicting actions leave their desires unclear. Although tribe members may have sought goods, "shooting arrows at people is a clear indication that they do not want contact." Official contact without a request is illegal, she says. She urges Peru to focus on keeping outsiders off the reserves.

Hill counters that "leaving the Mashco Piro alone will ultimately bring a terrible disaster to them," and calls the Peruvian government's intention to act "good news." But he warns that preventing epidemics among the group will require a commitment longer than the planned 6 months. Tribe members "cannot be left alone, even for a few weeks, during the first 2 to 3 years, or the whole population could go down," he says. Estremadoyro agrees, and worries that the government lacks the political and financial support to provide long-term food, shelter, medical help, and other services.

All agree that the situation is now a crisis. The government has "reacted far, far too late," Spooner says. She blames officials for failing to keep loggers, missionaries, and tourists off the Mashco Piro's land, which she says is at the root of the tribe's frequent appearances on the river. The government report suggests that outsiders are poaching animals such as peccary and tapir, which are staple foods for isolated people.

The ministry's August statement promises to patrol the river and train locals to avoid isolated people unless the tribe makes the first move. Assisting the Mashco Piro "is a huge challenge that cannot be postponed," Balbuena says. ■

Published by AAAS



www.sciencemag.org

THE LITTLEST FARMHANDS

Scientists are discovering thousands of microbes that help plants survive and thrive. Could these symbionts help farmers as well?

By **Jop de Vrieze** *in Palencia, Spain*

"Look. Nothing flourishes here. Not even weeds." Pius Floris picks up one of the dozens of stones scattered around him on the degraded, barren-looking soil. Decades of drought, monoculture, overuse of fertilizer, and excessive plowing have taken their toll on this field in the Spanish region of Castilla y León. As a result, wind and rain have washed away all but 25 centimeters of the fertile topsoil that used to nourish the grain here. For centuries, this area was a bread basket; today, yields are so low that farmers work the area only because of subsidies from the European Union.

Floris, a Dutch entrepreneur in plant health, wants to turn that situation around.

Mycorrhiza fungi (yellow) help the roots of this soybean plant absorb nutrients and water; to return the favor, the plant excretes nutrients for the fungi.

With researchers at the University of Valladolid and a team of local farmers, he participates in an E.U.-funded pilot project that aims to make profitable agriculture possible again on such damaged soil, without irriga-

PHOTO: MERTON BROWN/VISUALS UNLIMITED INC.

Published by AAAS

tion. His key collaborators: microbes.

One reason the soil here has gone to waste, Floris says, is that farming has destroyed its microbial ecosystem, which can help plants survive and thrive. His team has recently applied beneficial microbes—in particular fungi that live around plant roots—to this degraded area, in addition to organic fertilizer. "Farmers have ignored these symbionts for decades," Floris says. "We are bringing them back into the game."

He's not the only one trying to do that. Scientists have recently discovered a dizzying diversity of bacteria, viruses, and fungi that live in, on, and around plant roots in the soil. Many of these microbes, together called the rhizobiome, help plants one way or another, from providing nutrients to warding off crop pests and diseases. Small biotechnology and major plant science companies think they have huge potential benefits in agriculture and have recently begun a spate of new field trials. A 2013 report published by the American Academy of Microbiology (AAM), with the optimistic title *How Microbes Can Help Feed the World*, concluded that microbes have the potential to increase harvests while allowing farmers to use less fertilizer and pesticides; certain microbes can even enable plants to grow in very dry or salty places, which could help the world adapt to climate change.

A few denizens of the soil have been on the market for decades, such as *Trichoderma* fungi that suppress pathogenic fungi, and the now well-known caterpillar killer *Bacillus thuringiensis*, or *Bt*. (The gene for the bacterium's toxin has also been introduced into some crops' genomes.) Recently, major agrochemical companies such as Bayer have jumped on the biologicals bandwagon. "It's a revolution of microbiology," says Thomas Schäfer, vice president of microbial R&D at Novozymes, a company developing microbial fertilizers and pesticides that recently struck up an alliance with agriculture giant Monsanto. Schäfer believes farmers are headed for "precision agriculture," in which they would add beneficial microbes, or support existing ones, after an in-depth analysis of a field's microbiological makeup.

Whether these high expectations can be met by the budding research field is anyone's guess. The diversity in the soil microbiome is so staggering that finding out which organisms benefit plants most, how they do it, or what combinations work best is a gargantuan task. Also unclear is whether microbes can dramatically curb the use of pesticides and chemical fertilizers, and whether conventional farmers will trust these new options. The central question, some scientists say, is: How much can microbiology replace chemistry in agriculture?

SCIENTISTS AND FARMERS have long seen microbes primarily as problems. A fungus-like unicellular organism named *Phytophthora infestans*, responsible for potato blight and other crop diseases, has caused famines throughout history and is still a major problem. A variety of other fungi and bacteria cause the decay of roots and leaves.

To be sure, farmers know that some microbes are helpful: The group of bacteria called rhizobia, which live inside the roots of legumes and fix nitrogen from the air into a biologically useful form for the plants, are a textbook example. But recently, new methods of DNA sequencing and analysis have brought a vast, complex web of mutually beneficial interactions into view, comparable to the symbiotic roles researchers now believe are played by the thousands of bacterial and viral species that inhabit the human gut, skin, and other tissues.

Studies have shown that there are up to 10 billion bacterial cells per gram of soil in and around plant roots, a region known as the rhizosphere. This domain is tremendously diverse; in 2011, a team led by soil microbiologist Jos Raaijmakers of the Netherlands Institute of Ecology in Wageningen detected more than 33,000 bacterial and archaeal species on sugar beet roots. Dozens of species appear to suppress plant disease by excreting substances that ward off pathogenic microbes or occupying niches otherwise taken up by the pathogens. A study of vineyards in New York, published in March, showed that the composition of the rhizobiome depends heavily on the soil type.

Lab and greenhouse experiments have also shown that microbes make a variety of nutrients and minerals in the soil available to plants, produce hormones that spur growth, stimulate the plant immune system, and trigger or dampen its stress responses. "In general, we can say that a more diverse soil microbiome results in fewer plant diseases and higher yield," Raaijmakers says.

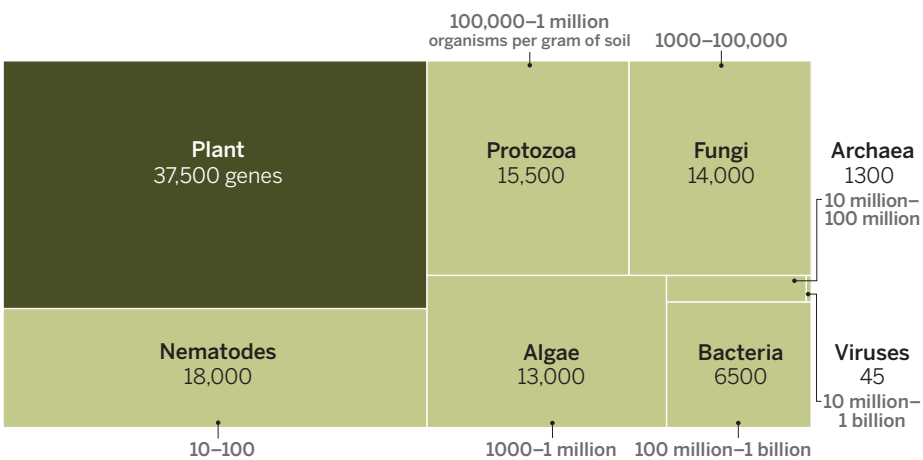
Among the most helpful microbes are so-called mycorrhizae or root fungi, which form a dense network of thin filaments reaching far into the soil, acting as extensions of the plant roots they live on or in. These fungi facilitate the uptake of water and a wide range of nutrients—Floris calls them "the plant's shopping carts."

Microbes can also help plants survive extreme conditions. A 2007 study showed that a complex symbiosis with fungi and viruses makes it possible for a grass called *Dichanthelium lanuginosum* to thrive in geothermal soils in Yellowstone National Park, where temperatures reach 60°C. The fungus, now thoroughly studied and introduced in the U.S. market in 2014 for application on corn and rice, triggers a stress response that the plants can't switch on themselves.

Similarly, a bacterium called *Stenotrophomonas rhizophila* has been shown to strongly increase drought tolerance in crops like sugar beets and maize. A 2013 study offered an explanation: The microbe excretes a variety of molecules that help plants withstand stress, including so-called osmoprotectants, which prevent the catastrophic outflux of water from plants in very salty environments. Microbes can even affect the flavor of food plants: A bacterium called *Methylobacterium extorquens* increases the production of furanones, a group of molecules that gives strawberries their characteristic flavor.

Underground world

Vast numbers of organisms—presented here along with their average number of genes and their numbers per gram of soil—live in and around a plant's root system.



GRAPHIC: ADAPTED FROM R. MENDES *ET AL.*, *FEMS MICORBIOLOGY REVIEWS* **37** (22 JULY 2013) © 2013 FEDERATION OF EUROPEAN MICROBIOLOGICAL SOCIETIES BY A. CUADRA/*SCIENCE*

Published by AAAS

The services provided by microbes are apparently hugely important to plants, as they put in a lot of energy to return favors. Studies have shown that up to 30% of the carbon fixed by plants is excreted from the roots as so-called exudates—including sugars, amino acids, flavonoids, aliphatic acids, and fatty acids—that attract and feed beneficial microbial species while repelling and killing harmful ones.

THE GROWING ACADEMIC understanding of the rhizobiome has increasingly made its way into the corporate world and onto farmers' fields. One early example was Serenade, a biopesticide containing a *Bacillus subtilis* strain that has antifungal and antibacterial properties and promotes plant growth. It was discovered by AgraQuest, a biotech in Davis, California. "So many pharmaceutical products were extracted from the soil, but for agriculture, this potential was hardly exploited," recalls Denise Manker, who cofounded the company in 1995. Serenade, registered by the U.S. Department of Agriculture in 2001, can be applied in a liquid form on the plants and in the soil to fight a range of pathogens. "Initially, most of our customers were organic farmers," Manker says. Soon, she says, innovative conventional farmers started experimenting with the product as well, and some became converts.

So far, the market for such products has been modest. Almost all of the registered ones are biopesticides; the AAM report estimated that they bring in about $1 billion annually, which pales compared with the global markets for chemical pesticides and fertilizer, estimated at $50 billion and $60 billion annually, respectively. But big agrochemical companies see the potential of microbial alternatives. "It took us 17 years to get the big companies interested, but we made it," Manker says: In 2012, German agro giant Bayer bought AgraQuest for $425 million. Manker became Bayer's director of global agronomic development of biologicals, a job that comes with a €10 million annual research budget. She's using it to field-test dozens of new fungi and bacteria to replace chemical pesticides or serve as biostimulants, which promote the health and growth of crops.

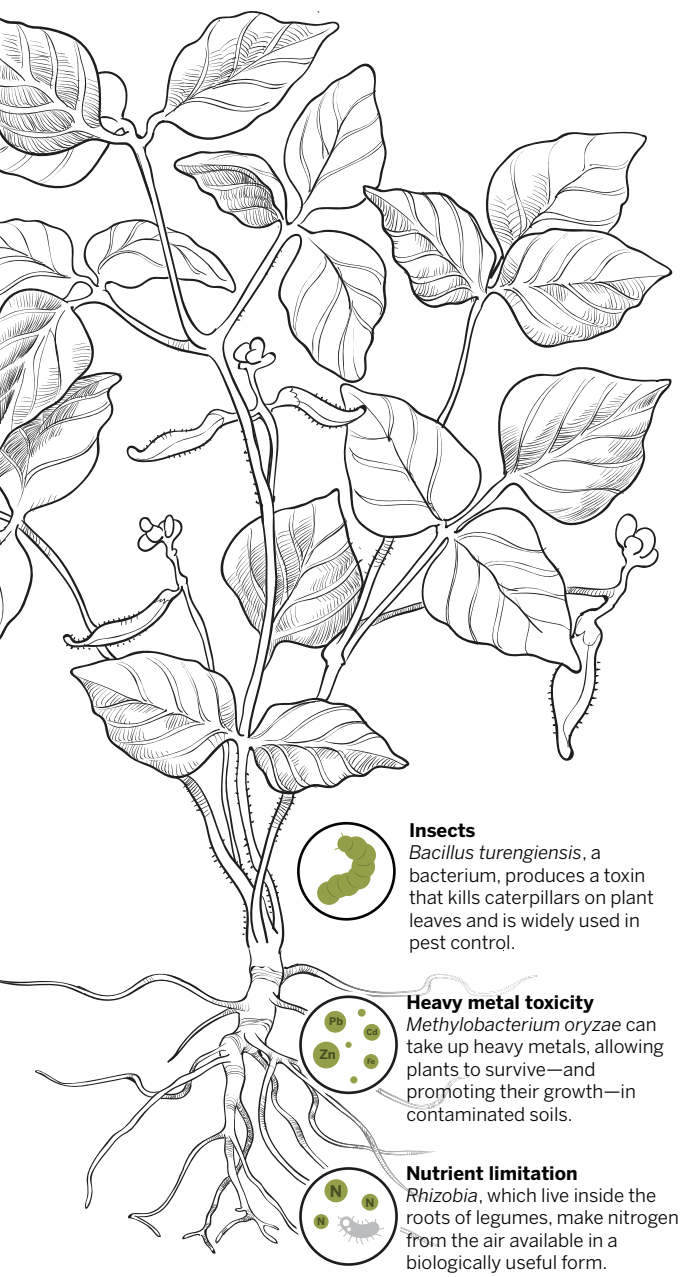
One explanation for Bayer's interest: Growing public resistance against chemical pesticides and a 2009 European directive aiming to reduce their use caused the market for chemical crop protection to stagnate, whereas the demand for biologicals was growing close to 10% per year. Given that, it's not surprising that Bayer's competitors have made similar moves. Syngenta and BASF acquired smaller companies developing microbial products last year; so did Dupont in April of this year. Monsanto's new partner, Novozymes, has invested heavily in a biofertilizer containing the soil fungus *Penicillium bilaii*, and a bioinsecticide that contains the fungus *Metarhizium anisopliae*.

The list of potentially suitable microbes is endless, says Matteo Lorito, a plant pathologist at the University of Naples Federico II in Italy, and that poses a daunting task for companies. "The challenge they are facing is selection of the ones that are commercially viable and effective," Lorito says, especially because many microbes are plant-specific and the composition of the rhizobiome can change rapidly.

Traditionally, selected microbes were first tested and investigated extensively in labs and greenhouses. But promising strains often failed to prove effective in the field, because of soil, climate, and ecosystem effects. Today, most companies use a "field-first approach," in which hundreds or even thousands of microbial strains are tested on field plots. If one proves successful, the mechanism of action is unraveled in the lab later. But even a promising field

How microbes help plants

Microorganisms living in roots and the soil can shield plants from a wide variety of threats. Here are some examples.



Heat
A symbiosis of *Curvularia* fungi and the *Curvularia* thermal tolerance virus living in the roots can increase plants' heat tolerance by more than 20°C.

Cold
A cold-tolerant *Pseudomonas* bacterium helps plants grow by fixing nitrogen from the atmosphere at temperatures as low as 4°C.

Drought
Mycorrhiza fungi such as *Glomus deserticola* extend the root system and can provide water from deeper soil layers to the plant.

Flooding
Enterobacter cloacae, a bacterium, protects plants from the destructive impact of flooding by reducing levels of a stress hormone that impairs root growth.

Osmotic stress
A bacterium called *Stenotrophomonas rhizophila* can excrete osmoprotectants, which prevent the catastrophic outflow of water from plants in very salty environments.

Pathogens
Soil fungi named *Trichoderma* can kill and outcompete pathogenic fungi and can activate the plant's own immune system.

Insects
Bacillus turengiensis, a bacterium, produces a toxin that kills caterpillars on plant leaves and is widely used in pest control.

Heavy metal toxicity
Methylobacterium oryzae can take up heavy metals, allowing plants to survive—and promoting their growth—in contaminated soils.

Nutrient limitation
Rhizobia, which live inside the roots of legumes, make nitrogen from the air available in a biologically useful form.

CREDITS: (ILLUSTRATION) RICHARD JOHNSON; (ICONS) C. SMITH/*SCIENCE*

Published by AAAS

study doesn't guarantee success on the farm. "People are using microbial products on a variety of crops, with different application methods and in different soils and climates," says Matthew Wallenstein, an ecosystem scientist at Colorado State University (CSU), Fort Collins. "That will make the results a lot more variable. It's very hard to make a miracle product that works everywhere."

That biologicals are living things is part of the problem; to work best, they need to become established and thrive on their own. One way to give them an edge is to apply them on plant seeds instead of into the soil; that way, they can enter the plant's rhizosphere early on as the first roots form and have a better shot at dominating the space.

Populations of beneficial microbes also dwindle over time. Spraying Serenade, for instance, results in a high *B. subtilis* density in the soil initially, but levels rapidly decrease during a farming season as the bacterium fails to obtain a permanent niche. That may be because it's outcompeted by the existing community of microbes. "Applying just that one strain is often not enough," Raaijmakers says. "You need a consortium of two, three, or even five or more collaborating strains that can withstand the ecological forces." To find successful combos, scientists have recently begun selecting these combinations in a systematic way, by identifying naturally occurring microbial networks in the field and studying their interactions down to the molecular level in the lab. "Scientifically, that's the way to go," Raaijmakers says.

Registering cocktails as biopesticides is a challenge, however, he says. Both in the United States and Europe, companies have to provide regulatory authorities with evidence that both the individual strains and the product as a whole are safe for consumers and the environment. "This is a laborious and expensive process," Raaijmakers says. That's why many of the existing products are not labeled "biopesticides," but "biostimulants"; the latter category is easier to get registered, but the market is less lucrative.

Because centuries of breeding may have robbed crop plants of an ability to attract beneficial microbes themselves, Raaijmakers also leads a project to study ancient crops and their microbes in their natural environment, such as wild beans in Colombia and the wheat ancestor *Triticum tauschii* (Tausch's goatgrass). The hope is to identify plant traits and symbiotic microbes that could benefit modern crops. The effort could yield compounds that can be applied to plants or soil, but in the long run, scientists hope to find the genes encoding the exudate molecules that attract microbes and reintroduce them into modern crops. "For the first time, breeders and biocontrol people are talking with each other," Lorito says. "They are now starting to include the interaction with microbes into the breeding."

FLORIS, THE DRIVING FORCE behind the pioneering project to restore the soil in Spain, has helped various factions speak with one another. Decades ago, while working for a Dutch tree care company—and studying with a famed plant pathologist of the U.S. Forest Service—he realized that roots do much more than anchor trees and suck up water. Today, he owns a company that offers advice and biological products to improve soil life to farmers, gardeners, and horticulturists; Raaijmakers and several other scientists also tap his practical knowledge.

Pius Florus is using a mix of microbes and organic pesticides to help restore a depleted field in Spain.

Floris's vision is more radical than most: a "cure" for degraded farmland, based on a holistic set of measures that puts microbes front and center. The hope is that this will even do away with the need for artificial fertilizer, "the mother of all evil," according to Floris, because it's disastrous for microbial life. Fertilizers deplete soil of organic matter and trace elements, cause salination, and suppress mycorrhizae; studies have shown that they can also turn symbiotic bacteria into competitors that kill each other.

The pilot project, now in its third year and as-yet unpublished, suggests that Floris's approach may work. The team borrowed an ancient Roman plow from a local museum to loosen the soil just slightly and made ridges to retain rainwater. They seeded oat plants along with vetch, which attracts bacteria that fix nitrogen and leave it in the vetch's roots after the harvest. They planted small olive trees to boost microbial diversity. Then they split the 100-hectare field into three zones. Zone A was treated with chemical fertilizer and pesticides; zones B and C were supplied with different amounts of an organic biofertilizer, consisting of fermented grape leftovers containing a variety of bacteria and fungi, and a dose of four different types of mycorrhiza spores. None of the areas were irrigated.

In May of this year, Floris strode through hip-high oats and vetch fields, bending over to inspect the crops every 2 meters. The crops in zone B, which received the most organic fertilizer, had reached nearly twice the height of those in zone A and were inches taller than zone C. "This really turns me on!" Floris shouted. (The yield of zone B, it would turn out after the harvest, equaled that of irrigated crops, whereas that of the conventional zone was negligible.)

Why this was possible became clearer when Floris's colleague Pedro Alonso dug a deep hole in zone B. Plant roots had found their way almost 2 meters into the rocky soil, deep enough to reach the groundwater. This could not have happened without the mycorrhiza, which penetrated the rock by excreting acids, Floris says: "These fungi enable the crops to survive without irrigation, even through the driest parts of the year."

Others are following the experiment with interest, but not everyone is convinced that Floris's recipe is the solution for agriculture as a whole. "If you want to make a difference, you need to fit into the existing system," CSU's Wallenstein says. "It is unrealistic to expect a sudden radical change. For farmers, it would just be too much of a financial risk to abandon fertilizers and pesticides and invest in new equipment to apply biologicals."

Prem Bindraban, executive director of the industry-backed Virtual Fertilizer Research Center in Washington D.C., acknowledges that chemical fertilizers can harm beneficial microbes. "But in my view, chemical fertilizer is still essential to maintain the yield," he says. "We need to find a solution to combine the benefits of fertilizer and symbiotic microbes." Novozymes's vice president Schäfer agrees. "Biologicals can help reduce the amount of fertilizer and water used, but there will still be a role for chemistry," Schäfer says.

But Floris wants to see just how far he can push his microbial helpers. On that May day, as he leaves the Spanish acre and returns to his car, a tractor passes in an adjacent field, leaving a cloud of organic dust. "Plowing, at 30°, midseason? A disaster for the soil and its inhabitants." He sighs. "Well, we shouldn't be too pessimistic. A lot of change is happening already, and more rapidly than ever." ■

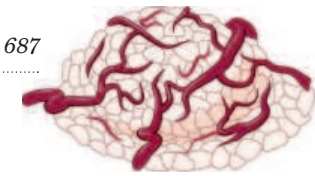
PHOTO: MIGUEL ÁNGEL SANTOS/AGENCIA GRÁFICA PHOTOGENIC

Published by AAAS

INSIGHTS

Mountain range formation p. 687

Rethinking vascular therapy for cancer p. 694 ▸



PERSPECTIVES

WATER

Watching water: From sky or stream?

Monitoring and management of freshwater resources has long depended upon on-the-ground measurements. Satellite remote sensing has brought new complementing capabilities. In this final of three debates, *Science* invited arguments about the appropriate roles for, and balance between, each approach.

Satellites provide the big picture

By **J. S. Famiglietti,[1,2,3]* A. Cazenave,[4,5] A. Eicker,[6] J. T. Reager,[1] M. Rodell,[7] I. Velicogna[1,2]**

Satellite observations have revolutionized our understanding of hydrology, water availability, and global change, while catalyzing modern advances in weather, flood, drought, and fire prediction in ways that would not have occurred with relatively sparse ground-based measurements alone. Earth-observing satellites provide the necessary "big-picture" spatial coverage, as well as the regional-to-global understanding essential for improving predictive models and informing policy-makers, resource managers, and the general public.

POLICY

Sustained investments in a robust satellite hydrology program have enabled a plethora of discoveries, along with modernization of water management, that have increased the human, economic, and water security of many nations. We now recognize distinct human- and climate-driven fingerprints on the water landscape that are dramatically changing the distribution of freshwater on Earth (*1*). Improved understanding and heightened societal awareness of the global extent of sea-level rise (*2*), ice sheet and glacial melt (*3*), changing rainfall patterns (*4*), declining snow cover (*5*), groundwater depletion (*6*), and the changing extremes of flooding (*7*) and drought (*8*) simply would not have occurred without satellite observations.

As we look ahead, ongoing and near-future missions will soon provide routine global monitoring of the stocks of soil moisture (*9*), surface water (*10*), and total water storage (*11*)—which will improve estimates of groundwater storage changes (*12*)—and of the fluxes of precipitation (*4*) and evapotranspiration (*13*). Taken to-

[1]*NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.* [2]*Department of Earth System Science, University of California, Irvine, CA, USA.* [3]*Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA.* [4]*Centre National d'Etudes Spatiales–Laboratoire d'Etudes Géophysique et Océanographique Spatiales (CNES/LEGOS), Toulouse, France.* [5]*International Space Science Institute, Bern, Switzerland.* [6]*Institute of Geodesy and Geoinformation, University of Bonn, Germany.* [7]*NASA Goddard Space Flight Center, Greenbelt, MD, USA.*
**Corresponding author. E-mail: james.famiglietti@jpl.nasa.gov.*

ILLUSTRATION: DAVIDE BONAZZI

Published by AAAS

gether, these measurements will enable improved characterization of terrestrial, atmospheric, cryospheric, and oceanic water budgets at multiple scales and will allow for evaluation of their responses to climatic variability (*14*). Small-scale airborne measurements of snow-water equivalent are paving the way for a global satellite mission (*15*). Space-geodetic measurements—e.g., from global positioning system (GPS) and interferometric synthetic aperture radar (InSAR)—can measure Earth's crustal response to the addition and removal of water. Such measurements are proving useful in tracking the dynamics of terrestrial water storage (*16*).

TRANSCEND POLITICS, INFORM POLICY. An important feature of satellite observations is that measurements are readily collected and shared across political boundaries. In contrast, many countries refuse to share ground-based hydrologic measurements for socioeconomic and political reasons, whereas others simply lack the capacity to centralize and digitize what data are collected. This impedes investigations on a continental-to-global scale, as well as efforts to avert food and water crises. Because satellites can monitor water resources at scales relevant to effective transboundary water management and because data are often provided through freely accessible digital archives, policies of international data denial may ultimately become obsolete (*17*).

Governments around the world are now instructing their water management agencies to plan for the uncertain hydrologic future that satellite observations have helped reveal. The United States recently ordered its Bureau of Reclamation to rethink water storage strategies to better respond to prolonged drought and climate change impacts in its arid west. India has undertaken a national hydrogeological mapping program to better characterize its available groundwater resources. Australia, Israel, and several other countries have a long history of adaptive water management, guided in part by satellite observations.

Satellite-based studies, such as those of California drought and groundwater depletion (*18*), have affected water policy, e.g., the passage of California's Sustainable Groundwater Management Act of 2014. California will finally manage its once-vast groundwater supply to prolong its availability for future generations. Satellite observations proved highly informative for elected officials, policymakers, and the public.

PRIORITIZING INVESTMENTS. In an era of increased competition for limited federal funding, investments in satellite hydrologic monitoring should be critically evaluated for their anticipated returns, compared with investments in other technologies, including ground-based measurements. Several of us debated the relative merits of the Surface Water and Ocean Topography (SWOT) mission (*10*) to measure the heights, slopes, and inundated areas of water in large rivers and other inland bodies. Would a similar expenditure in traditional stream gauging provide equal science, technology, and management returns on investment? Arguments in favor of the satellite approach prevailed, the SWOT mission earned broad support from the water-science community, and it was ultimately selected for funding by NASA and Centre National d'Etudes Spatiales, with launch scheduled for 2020.

Such support is not always available nor should it be. We fully support comprehensive, ground-based measurements as the backbone of a regional-to-global hydrologic observing network. In situ observations provide important validation data for satellite measurements and are typically collected with greater spatial and temporal frequency. Some measurements, such as the volume of groundwater stored in major aquifers, are poorly suited to remote observation and may only be made with measurements acquired on land (*19*). There is no substitute for a well-maintained in situ network, and we lament the decline of such networks around the world.

Additional investment in data-model integration could help maximize the utility of current and forthcoming satellite hydrology missions. Data-model integration platforms are likely the most reliable means for quantifying freshwater availability at regional scales, as well as for down-scaling coarser-resolution satellite observations to the finer-resolution scales at which regional predictions and water management decisions are made. High-resolution models that represent the main components of natural and managed water cycles and that can ingest ground-, aircraft-, and satellite-based observations should have accelerated development timelines (*20*).

Satellites play a central role in scientific and operational hydrology and water management. Reliable hydrometeorological prediction would not be possible without them nor would rapid response to emergencies like regional flooding. With the scientific community recognizing that the water cycle is changing in profound ways (*1*), satellites provide the best available means to characterize these changes over large regions, to better understand and predict their implications for humanity, and to communicate compelling findings to elected officials and environmental decision-makers. ■

REFERENCES

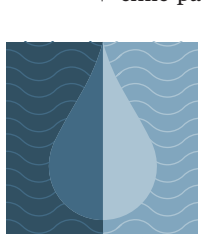
1. J. S. Famiglietti, M. Rodell, *Science* **340**, 1300 (2013).
2. A. Cazenave, W. Llovel, *Annu. Rev. Mar. Sci.* **2**, 145 (2010).
3. E. Rignot, J. Mouginot, B. Scheuchl, *Geophys. Res. Lett.* **38**, L05503 (2011).
4. A. Hou *et al.*, *Bull. Am. Meteorol. Soc.* **95**, 701 (2014).
5. S. J. Déry, R. D. Brown, *Geophys. Res. Lett.* **34**, L22504 (2007).
6. A. S. Richey *et al.*, *Water Resour. Res.* 10.1002/2015WR017349 (2015).
7. J. T. Reager, J. S. Famiglietti, *Geophys. Res. Lett.* **36**, L23402 (2009).
8. A. Thomas, J. T. Reager, J. S. Famiglietti, M. Rodell, *Geophys. Res. Lett.* **41**, 1537 (2014).
9. D. Entekhabi *et al.*, *Proc. IEEE* **98**, 704 (2010).
10. M. Durand *et al.*, *Proc. IEEE* **98**, 766 (2010).
11. B. D. Tapley, S. Bettadpur, J. C. Ries, P. F. Thompson, M. M. Watkins, *Science* **305**, 503 (2004).
12. M. Rodell, I. Velicogna, J. S. Famiglietti, *Nature* **460**, 999 (2009).
13. M. C. Anderson *et al.*, *Hydrol. Earth Syst. Sci.* **15**, 223 (2011).
14. M. Rodell *et al.*, *J. Clim.* 10.1175/JCLI-D-14-00555. (2015).
15. J. S. Deems, T. H. Painter, J. J. Barsugli, J. Belnap, B. Udall, *Hydrol. Earth Syst. Sci.* **17**, 4401 (2013).
16. A. A. Borsa, D. C. Agnew, D. R. Cayan, *Science* **345**, 1587 (2014).
17. J. S. Famiglietti, *Nat. Clim. Change* **4**, 945 (2014).
18. J. S. Famiglietti *et al.*, *Geophys. Res. Lett.* **38**, L03403 (2011).
19. A. S. Richey *et al.*, *Water Resour. Res.* 10.1002/2015WR017351 (2015).
20. E. F. Wood *et al.*, *Water Resour. Res.* **47**, W05301 (2011).

10.1126/science.aac9238

Time for in situ renaissance

By **Balázs M. Fekete,[1]* Richard D. Robarts,[2] Michio Kumagai,[3] Hans-Peter Nachtnebel,[4] Eric Odada,[5] Alexander V. Zhulidov[6]**

In situ monitoring of water dates to Pharaonic Egypt and remained the primary means of observation into the later part of the 20th century. Monitoring networks have declined (*1–4*) since the 1980s because of budgetary constraints and political instabilities. This decline paradoxically has coincided with growing interest in climate change. The rise of satellite remote sensing promised global observing capabilities and put in situ monitoring on the sidelines. Capabilities offered by in situ monitoring versus satellite remote sensing are very different and mostly complementary (*5*); thus, deployment should depend on monitoring requirements (observed parameter, data quality, spatiotemporal scale, data costs, and access).



WATER DEBATE SERIES

Monitoring systems in situ support water manage-

Published by AAAS

ment and policy development, as well as serving a range of users and uses (e.g., agricultural operations, environmental management, and regional planning). Remote sensing depends on in situ monitoring for essential calibration and validation. Water managers tend to use in situ observations because they need continuous, long-term, high-frequency, and accurate data for designing infrastructures and effective management plans, as well as sustained real-time data for operation.

FIDELITY, RESOLUTION, CONSISTENCY. Only in situ sensors, typically in close contact with the monitored medium, can measure a host of water-related quantity and quality parameters and processes (*6*) with reliable accuracy and sufficient frequency. Remote sensing provides indirect measurements normally limited to the near surface of the monitored object and affected by the media between the sensors and the monitored object. Remote-sensing observations are often the result of complex retrieval algorithms. In extreme cases, like satellite-derived evapotranspiration (*7*–*9*), the algorithm is almost indistinguishable from land surface hydrology models, such that it is questionable that this qualifies as "observation."

In situ observations are better suited for gradually changing observational targets, when strategically placed point measurement sensors are representative for larger areas. River discharge in particular is an ideal target for point monitoring because discharge only changes gradually along a river channel (except for confluences) and represents an integrated signal of the hydrological processes from a larger area upstream (*1*). Unless measurement requires laboratory processing of samples, in situ monitoring can provide observations at high temporal frequency. Many in situ observational records cover multiple decades of continuous data at high temporal resolution. Observation consistency depends on continuous instrument maintenance and recalibration that is often the most expensive part of the monitoring program. Remote sensing that only replaces relatively inexpensive measurements without comparably rigorous calibration will compromise monitoring (*5*).

Satellites are placed either in geostationary orbit, where they can provide continuous observations at low spatial resolution, or in low Earth orbits, which results in low repeat frequencies flying over the same area unless a constellation of satellites is deployed at added expense. It can be difficult to derive continuous (multidecadal) time series from satellite records, because technology changes and space agencies do not pay adequate attention to the homogeneity of observational records. Many satellite platforms (with the exception of meteorological satellites in geostationary orbits) are still in an "experimental" phase without long-term commitment for continued operations. Satellite sensors without adequate backup present a single point of failure leading to abrupt termination of observations.

COST, INNOVATION, ACCESS. Cost comparison of satellite remote sensing versus in situ monitoring is difficult because the final products are rarely comparable. Satellite remote sensing only competes in large-scale or global applications, because it cannot replace in situ monitoring in most cases. Cost comparison should be posed as the additional expense of extending existing in situ monitoring, including incentives for data sharing and aggregating observations, versus operating an independent satellite monitoring infrastructure. A recent World Bank report (*10*) estimated that $1.5 to $2 billion would be necessary to modernize developing countries' hydrometeorological monitoring infrastructure and an additional $0.4 to $0.5 billion annually for maintenance. These are comparable to the typical $0.3 to $0.6 billion price tag of medium-sized satellite missions.

Telecommunication breakthroughs and their widespread use lower barriers to data transmission. New sensor and deployment technologies are improving performance and cost. Autonomous drone vehicles (aircraft, boats, or submarines) could operate as monitoring platforms, which would blur the distinction between remote sensing and in situ observations. Solar unmanned aerial vehicles may offer cost-effective alternatives to satellites.

> ***"[D]ata sharing is voluntary, and agencies collecting in situ observations rarely have the obligation or incentives to share."***

Differences in in situ monitoring are inevitable to meet specific needs, but much monitoring could be standardized to ease data processing over larger geographic domains. Spatiotemporal synthesis of in situ observations often leads to more refined and accurate assessment. Yet lack of international collaboration in data sharing is often a motivation to develop remote-sensing alternatives. Global data centers are criticized for inefficiencies in collecting and disseminating in situ data. But data sharing is voluntary, and agencies collecting in situ observations rarely have the obligation or incentives to share. International agreements [e.g., the Danube, Mekong, Zadec, Rhine basins; World Meteorological Organization (WMO) Resolution 25 (*11*)] and data centers [GRDC, GPCC, and GEMS/Water (*12*)] are paving the way to improved standardization and access for in situ monitoring data. Investments in in situ monitoring and data centers at funding levels comparable to satellite remote sensing, contingent upon unrestricted access to data, likely can break many data-sharing barriers.

Sustained coordination and maintenance of in situ observing networks is far more challenging than flying a few satellites but could improve the quality of observations and serve as a positive precedent for international collaborations that fosters trust among nations. Succeeding in coordinated efforts for improved Earth observations could encourage commitments to larger goals like combating climate change. ■

[1]CUNY Environmental CrossRoads Initiative, Department of Civil Engineering, The City College of New York, City University of New York, New York, NY 10031, USA. [2]World Water and Climate Foundation, Saskatoon, Saskatchewan S7N 4A2, Canada. [3]Research Organization of Science and Technology, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan. [4]Institute of Water Management, Hydrology and Hydraulic Engineering, University of Natural Resources and Life Sciences, 1190 Vienna, Austria. [5]College of Physical and Biological Sciences, University of Nairobi, Nairobi, Kenya. [6]South Russian Centre for Preparation and Implementation of International Projects, 344090 Rostov-on-Don, Russia.
*Coresponding author. E-mail: bfekete@ccny.cuny.edu

REFERENCES AND NOTES

1. E. Stokstad, *Science* **285**, 1199–1200 (1999).
2. A. V. Zhulidov, V. V. Khlobystov, R. D. Robarts, D. F. Pavlov, *Can. J. Fish. Aquat. Sci.* **57**, 1932–1939 (2000).
3. C. J. Vörösmarty *et al.*, *Eos Trans.* **82**, 54 (2001).
4. A. I. Shiklomanov, R. B. Lammers, C. J. Vörösmarty, *Eos Trans.* **83**, 13 (2002).
5. B. M. Fekete, U. Looser, A. Pietroniro, R. D. Robarts, *J. Hydrometeorol.* **13**, 1977–1986 (2012).
6. R. D. Robarts, S. J. Barker, S. Evans, in *Proceedings of Taal 2007: The 12th World Lake Conference*, M. Sengupta and R. Dalwani, Eds., Jaipur, India (International Lake Environment Committee, Shiga, Japan, 2008), pp. 167–175.
7. W. G. M. Bastiaanssen, M. Menenti, R. A. Feddes, A. A. M. Holtslag, *J. Hydrol. (Amsterdam)* **212-213**, 198–212 (1998).
8. Q. Mu, F. A. Heinsch, M. Zhao, S. W. Running, *Remote Sens. Environ.* **111**, 519–536 (2007).
9. Q. Mu, M. Zhao, S. W. Running, *Remote Sens. Environ.* **115**, 1781–1800 (2011).
10. D. Rogers, V. Tsirkunov, *Weather and Climate Resilience: Effective Preparedness Through National Meteorological and Hydrological Services* (The World Bank, Washington, DC, 2013); www.worldbank.org.
11. WMO, *Exchange of Hydrological Data and Products* (WMO, Geneva, 1999).
12. The Global Runoff Data Center (GRDC), Global Precipitation Climatology Centre (GPCC), and United Nations Global Environment Monitoring System (GEMS/Water) provide long-term hydrological data.

10.1126/science.aac7358

Published by AAAS

GEOPHYSICS

How mountains get made

Observations of crustal deformation constrain models of mountain formation

By **Maureen D. Long**

The formation of mountain belts (orogens) in subduction-collision settings, where an oceanic plate subducts beneath continental material, is a fundamental process in plate tectonics. However, the mechanisms by which the continental crust deforms to produce shortening and uplift, and thus high topography, has been a matter of debate. This uncertainty is largely due to the difficulty of making direct observations of deformation in the deep crust to test the predictions made by conceptual models. On page 720 of this issue, Huang *et al.* (*1*) use observations of seismic anisotropy to constrain the geometry of deformation in the continental crust beneath the Taiwan orogen, and thus shed light on how the crust deforms as mountains are formed.

Two general concepts of crustal deformation in collisional orogens have been proposed, known as the thin-skinned and thick-skinned models (*2*–*4*). In the thin-skinned model, deformation is accommodated mainly in the upper crust, with a mechanically weak detachment surface (a décollement) separating the deforming upper layers from the deeper crustal rocks. In this scenario, the horizontal shortening and uplift required to form the mountains are confined to the upper crust. In contrast, the thick-skinned model invokes the deformation of the deeper crust as well as its shallow portions; here, the basement crustal rocks (as well as, perhaps, the mantle lithosphere beneath) undergo appreciable deformation. Taiwan represents an excellent locality to test these conceptual models (*5*); it is a young, actively deforming collisional orogen that accommodates the ongoing convergence between the Eurasian and Philippine Sea plates (see the figure).

A key challenge in discriminating among the different models of crustal deformation is the difficulty of constraining deformation in the deep crust. One type of observation that can shed light on deformation in the deep Earth is the characterization of seismic anisotropy, or the directional dependence of seismic wave speeds (*6*). In many regions of the Earth, including much of the crust, there is a relationship between strain and the resulting anisotropy: As a rock deforms, individual mineral crystals tend to rotate and form a statistical preferred alignment, giving rise to seismic anisotropy. Therefore, observations of anisotropy in the crust can constrain the depth distribution of collision-related deformation in orogens. Detailed observations of crustal anisotropy (and its variation in three dimensions) can be difficult, but recent innovations in observational seismology have advanced its study. These include the increasing availability of dense networks of seismometers, such as the Taiwan Integrated Geodynamics Research (TAIGER) deployment in Taiwan (*7*), and the use of the ambient seismic noise field to extract information about crustal structure, including its anisotropy (*8*).

Huang *et al.* develop a tomographic model of shear wave velocity and anisotropy in the crust beneath Taiwan using measurements of surface waves derived from ambient noise. They find evidence for a sharp change in the geometry of seismic anisotropy at a depth of around 10 to 15 km in the crust. Above this transition region, the fast directions of anisotropy are roughly parallel to the strike of the Taiwan orogen, and correlate closely with surface geologic trends. The authors propose that anisotropy in this upper layer is induced by compressional deformation and shortening. Beneath the transition, the fast directions of anisotropy are roughly parallel to the direction of convergence between the Eurasian and Philippine Sea plates. Here, the authors hypothesize that the deeper layer of anisotropy is caused by shear deformation of anisotropic minerals in the deep crust, induced by the motion of the downgoing plate.

As to what the results tell us about crustal deformation in the Taiwan orogen, and which of the thin-skinned or thick-skinned concepts apply, Huang *et al.* propose a hybrid model that has aspects of both

CREDIT: JEFF SCHMALTZ/MODIS RAPID RESPONSE TEAM, NASA GODDARD SPACE FLIGHT CENTER

Mountain-forming islands. The island of Taiwan, shown here in a radar image, sits off of the coast of southern China at the convergent boundary between the Philippine Sea plate and Eurasion plate. Huang *et al*. argue that Taiwan's mountainous terrain results from a mix of crustal deformation styles.

Department of Geology and Geophysics, Yale University, New Haven, CT, USA.
E-mail: maureen.long@yale.edu

Published by AAAS

concepts. The presence of strong anisotropy in the lower crust induced by convergence-parallel shearing implies that there is deformation throughout the crust, as suggested by the thick-skinned model. On the other hand, the evidence for a sharp change in deformation geometry at a depth of ~10 to 15 km shares some aspects of the thin-skinned concept. The authors propose that this transition is not a décollement in the traditional sense, as there is mechanical coupling between the upper and lower crustal layers; however, their model does imply that compressional tectonics is active only in the upper layer.

These results reported by Huang *et al.* have important implications for our understanding of how the crust deforms in collisional orogens, and may prompt a reexamination of other mountain belts. A key question is whether there is widespread lower crustal anisotropy in other orogens, and whether a transition in deformation style in the mid-crust is a universal feature. The implications of such a sharp transition for our understanding of crustal rheology need to be explored. Another important question is to what extent the mantle lithosphere, in addition to the lower crust, participates in deformation. More generally, the observation and interpretation of crustal anisotropy, both in mountain belts and in other tectonic settings, represents an exciting frontier area, enabled by the increasing availability of data from dense seismic networks and the maturation of observational techniques that rely on the ambient noise field (*1*, *8*) or on the analysis of converted waves (*9*). Furthermore, new constraints on the relationships between strain and anisotropy in crustal rocks (*10*, *11*) are enhancing our ability to relate seismic observations to deformation geometry, opening the door to the detailed and quantitative testing of hypotheses related to the deformation of Earth's crust. ■

REFERENCES

1. T.-Y. Huang, Y. Gung, B.-Y. Kuo, L.-Y. Chiao, Y.-N. Chen, *Science* **349**, 720 (2015).
2. D. Davis, J. Suppe, F. A. Dahlen, *J. Geophys. Res.* **88** (B2), 1153 (1983).
3. M. P. Coward, *J. Struct. Geol.* **5**, 113 (1983).
4. O. A. Pfiffner, *Spec. Pap. Geol. Soc. Am.* **414**, 153 (2006).
5. F. T. Wu, R.-J. Rau, D. Salzberg, *Tectonophys.* **274**, 191 (1997).
6. M. D. Long, T. W. Becker, *Earth Planet. Sci. Lett.* **297**, 341 (2010).
7. F. T. Wu, H. Kuo-Chen, K. D. McIntosh, *J. Asian Earth Sci.* **90**, 173 (2014).
8. M. P. Moschetti, M. H. Ritzwoller, F. Lin, Y. Yang, *Nature* **464**, 885 (2010).
9. V. Schulte-Pelkum, K. H. Mahan, *Earth Planet. Sci. Lett.* **402**, 221 (2014).
10. D. Ward, K. Mahan, V. Schulte-Pelkum, *Geophys. J. Int.* **190**, 1123 (2012).
11. B. Ko, H. Jung, *Nat. Commun.* **6**, 6586 (2015).

10.1126/science.aac8671

DEVELOPMENT

Marmoset kids actually listen

Humans and other vertebrates may share a developmental program for vocal learning

By **Daniel Margoliash**[1*] *and* **Ofer Tchernichovski**[2]

Undergraduate linguistics courses typically present language as unique to humans. Chomsky and others have postulated a language organ that evolved in hominids. This idea found modest support in the lack of evidence for vocal production learning (imitating sounds) in nonhuman primates. But did language suddenly emerge in the *Homo* lineage as a "hopeful monster" (*1*) who could learn new sounds and meanings? Evidence for vocal learning in nonhuman primates is now emerging (*2*, *3*), and in hindsight, looking at vocal production learning as the sole evolutionary precursor of language might have been shortsighted. Similar developmental processes can lead to different end points, and minor modifications of a primitive developmental program can create very different creatures. On page 734 of this issue, Takahashi *et al.* (*4*) provide evidence for a developmental process, rather than its endpoint, which reveals a shared developmental program for animal communication and human language. This indicates an ancestral developmental program that is shared not only between humans and other primates but also across mammals and birds.

Vocal imitation was long thought to be common in birds but rare in mammals. In the past few decades, scientists have reported evidence for vocal production learning in marine mammals, bats, elephants, and primates (*2*). It is difficult to work with large mammals that breed slowly and require social interactions for normal development, which precludes their being reared in isolation. These difficulties are particularly acute for primate research. Nevertheless, recent results from field studies and in captivity demonstrate shared, learned group signatures in communication calls produced by monkeys and apes. For example, when two troops of adult chimpanzees were integrated in a zoo—an event that is improbable in nature—the referential food grunts of introduced members apparently changed to match those of the home territory group (*3*).

Given that there may be no evolutionary canyon separating human vocal learning from that of other primates, one can attempt to compare vocal development across species that vary strongly in their end points. Takahashi *et al.* did just that. They thoroughly investigated vocal development in common marmosets (*Callithrix jacchus*), tiny social primates that produce a limited number of call types. In captivity, their calls are dominated by whistlelike calls (phees). The authors report rapid changes in vocal development driven by social (parent-infant) interactions. These vocal changes are similar to those previously described in human infants (*5*) and songbirds (*6*).

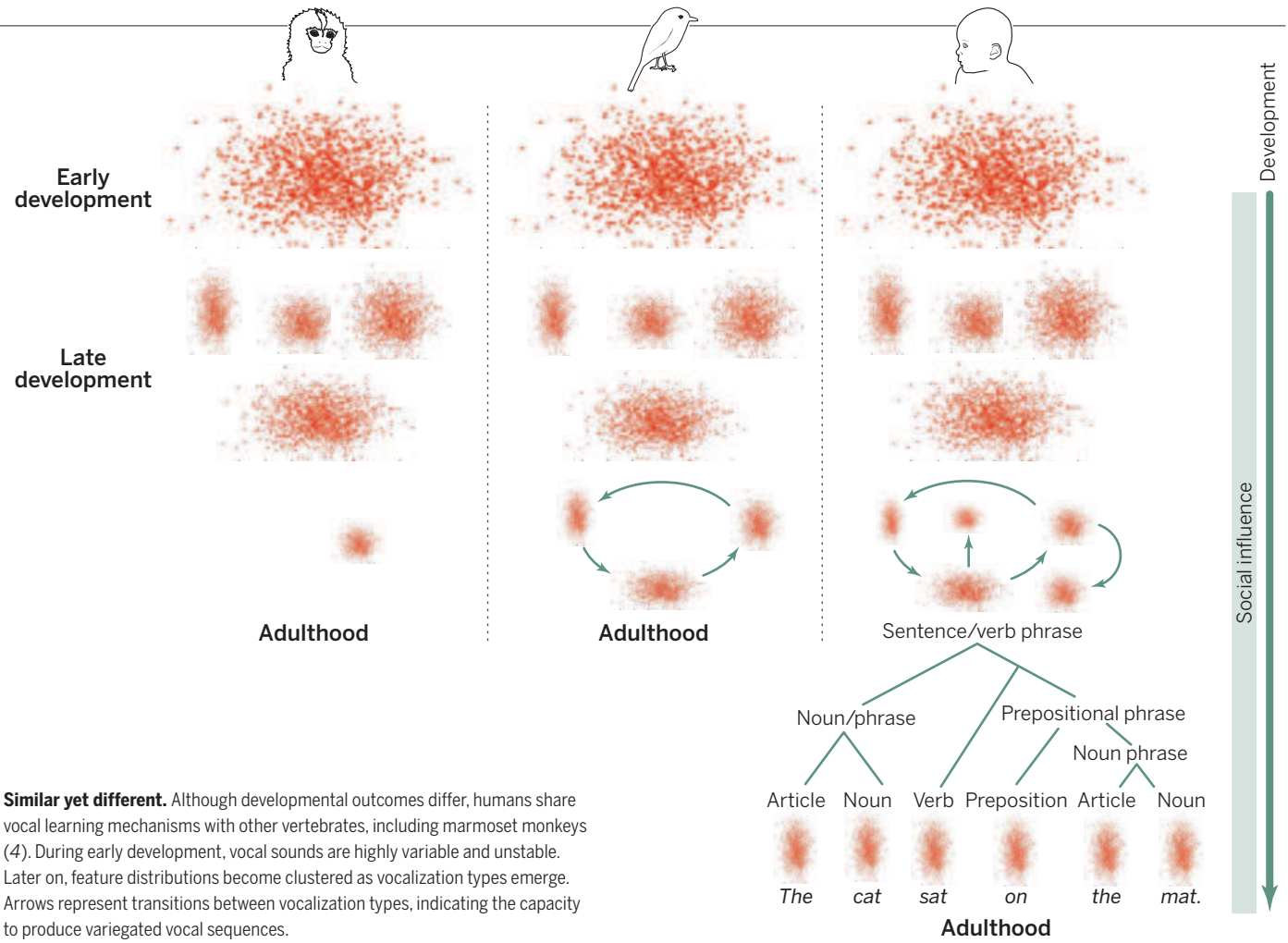
> ***"... Takahashi et al.'s findings point to an ancient substrate for vocal learning that an evolving large hominid brain could take advantage of ..."***

Takahashi *et al.* studied marmoset infants for 2 months from birth. By the end of this period, the juveniles were producing adult-like calls. Using a design that allowed dense sampling of thousands of vocalizations, the authors analyzed developmental changes in acoustic features previously shown to be useful for characterizing birdsong development. They observed significant vocal changes during development (see the figure). Early on, vocalizations were highly variable. Later, acoustic features of calls became clustered, indicating the emergence of ethologically meaningful call types. The numbers of clusters then declined, resulting in part from the transformation of immature call types into mature versions of phees.

To model the marmoset vocalizations, the authors used a nonlinear dynamical systems model of the vocal tract based on shared physical properties of vocal production in birds and mammals (*7*). Different calls occupied different regions of model parameters, which suggested that the changes are not ac-

[1]*Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA.* [2]*Department of Psychology, Hunter College, City University of New York, New York, NY 10065, USA.* **E-mail: dan@bigbird.uchicago.edu*

Published by AAAS



Similar yet different. Although developmental outcomes differ, humans share vocal learning mechanisms with other vertebrates, including marmoset monkeys (*4*). During early development, vocal sounds are highly variable and unstable. Later on, feature distributions become clustered as vocalization types emerge. Arrows represent transitions between vocalization types, indicating the capacity to produce variegated vocal sequences.

counted for by growth alone. The biophysical model coupled with electromyographic measures of respiratory effort show that the marmoset infant cries, subharmonic phees, and mature phees are developmentally related. Infant cries arise from low respiratory pressure and laryngeal muscle tension. In contrast, adult phees are characterized by high pressure and tension. Interestingly, subharmonic phees occupy an intermediate regime.

These extensive analyses set the stage to address the critical question: Can infant-parent interactions help to explain the developmental trajectory of marmoset vocalizations? Takahashi *et al.* observed that infants rapidly transitioned to producing more adult than infant vocalizations. The timing of that transition varied considerably across individuals and was strongly correlated with the number of parental responses that each infant experienced.

How can we relate these behavioral results to an evolutionary process? Perhaps, just as evolution can be understood as a modification of a developmental program, we could think about vocal learning as a modification of a program for vocal development. The early stages of vocal development are remarkably similar across taxa (see the figure). The infant produces highly diverse but loosely structured vocalizations, a cloud of sounds from which distinct clusters gradually emerge. This indicates a transition from a continuous, graded signal to a weakly symbolic vocal performance. Call types then undergo further differentiation and selective attrition. A process for combinatorial capacity emerges.

The relative intensities of these processes vary across species. In the marmoset, there is modest attrition and differentiation of calls and little combinatorial capacity. In zebra finches, attrition and differentiation of syllable (learned-call) types are fairly balanced, which results in several syllable types and a moderate combinatorial capacity. In human infants, the differentiation of vocal types and combinatorial abilities are both extensive and prolonged and eventually lead to spoken language. From the perspective of neurobiology, the brainstem-midbrain systems for call production are common in vertebrates (*8*). Species-specific differentiation of vocal behavior might have evolved through gradual increase in the interactions between those primitive brain structures and the forebrain, driving species-specific variability in symbolic and combinatorial processes.

It has long been known that monkeys can learn during development to associate different vocalizations with different meanings (*9*). Takahashi *et al.* show that contingent social feedback also shapes the transition to adult vocal patterns, adding to the vocal learning processes that likely preceded the emergence of language. A single explanation for the complex factors influencing changes in vocal developmental patterns over evolutionary time is unlikely to emerge. However, Takahashi *et al.*'s findings point to an ancient substrate for vocal learning that an evolving large hominid brain could take advantage of, thus continuing the evolutionary process that has enabled communication in other animals (*10*). ■

REFERENCES

1. R. Dawkins, *Climbing Mount Impossible* (Norton, New York/London, 1996).
2. S. E. R. Egnor, M. D. Hauser, *Trends Neurosci.* **27**, 649 (2004).
3. S. K. Watson *et al.*, *Curr. Biol.* **25**, 495 (2015).
4. D. Y. Takahashi *et al.*, *Science* **349**, 734 (2015).
5. M. H. Goldstein, J. A. Schwade, *Psychol. Sci.* **19**, 515 (2008).
6. O. Tchernichovski *et al.*, *Science* **291**, 2564 (2001).
7. A. Amador *et al.*, *Nature* **495**, 59 (2013).
8. A. H. Bass, E. H. Gilland, R. Baker, *Science* **321**, 417 (2008).
9. R. M. Seyfarth *et al.*, *Science* **210**, 801 (1980).
10. D. Margoliash, H. C. Nusbaum, *Trends Cogn. Sci.* **13**, 505 (2009).

10.1126/science.aac7860

Published by AAAS

NEUROSCIENCE

Astrocytes tell neurons when to listen up

Astrocytes ensure communication between neurons of the same circuit

By Aryn H. Gittis[1,2] and Daniel J. Brasier[1]

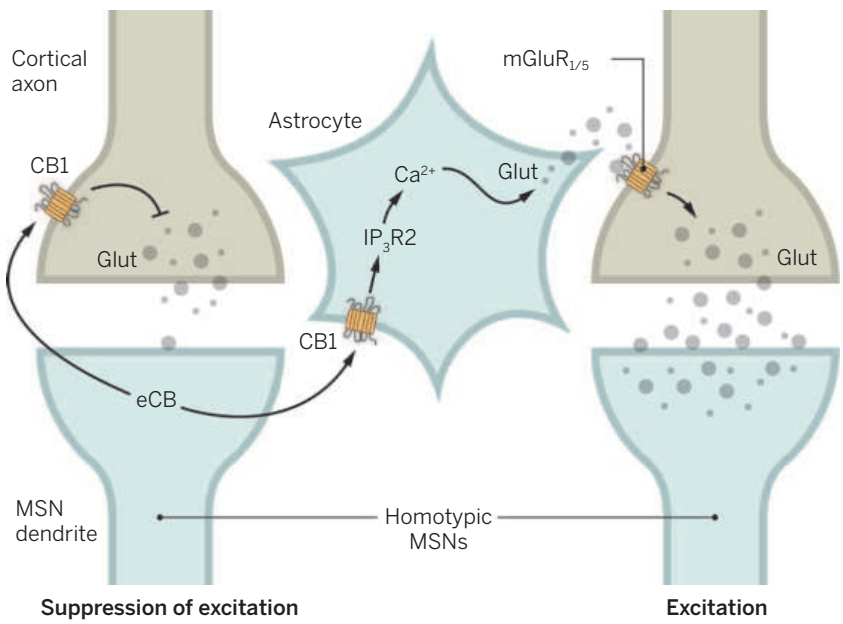
Throughout our nervous systems, neurons are faced with an enormous number of inputs, only some of which are relevant for accurate perceptions and appropriate behavior. On page 730 of this issue, Martín *et al.* (*1*) uncover a mechanism through which activation of one neuron temporarily enhances the responsiveness of neighboring neurons of the same type. This allows specific populations of neurons to effectively compare notes about their inputs, increasing the chance that if a group of neurons is activated simultaneously, a signal will propagate and change behavior. It turns out that the cellular machinery required to orchestrate this coordination extends beyond the neurons themselves to the surrounding network of glia, revealing a previously unrealized dimension of cellular specificity within the nervous system.

Glia, specifically astrocytes, are active partners in neuronal communication and plasticity of circuits (*2*). However, despite evidence of molecular heterogeneity in astrocytes (*3*), it had been unclear whether distinct populations of astrocytes could couple selectively to specific populations of neurons. With their discovery that cell-specific astrocyte-neuron networks exist in the striatum, Martín *et al.* raise the possibility that astrocytes selectively facilitate communication between neuronal ensembles in many other brain regions as well.

The specificity of astrocyte-neuron networks was demonstrated by Martín *et al.* in the striatum, a basal ganglia nucleus composed of anatomically intermixed, but functionally distinct, populations of projection neurons, called medium spiny neurons (MSNs), that contain the neurotransmitter γ-aminobutyric acid (GABA). When activated, MSNs expressing D1-type dopamine receptors (D1 MSNs) promote actions by disinhibiting the motor system through a circuit called the "direct" pathway. By contrast, when MSNs expressing D2-type dopamine receptors (D2 MSNs) become activated, they cancel actions by suppressing the motor system through a circuit called the "indirect" pathway (*4*). Because MSNs rely on excitatory inputs (the neurotransmitter glutamate) to become active, these synapses are important sites of regulation for behavioral control. Pathway-specific plasticity at glutamatergic synapses promotes learning (*5*), but can also contribute to motor dysfunction in disease (*6*).

Taking advantage of transgenic mouse lines to fluorescently identify D1 and D2 MSNs, Martín *et al.* observed that activation of an MSN in one pathway reduced excitation onto that neuron through a well-characterized endocannabinoid-mediated mechanism [depolarization-induced suppression of excitation (DSE) (*7*)], but unexpectedly, simultaneously enhanced excitatory inputs onto neighboring neurons in the same pathway (see the figure). This heterosynaptic potentiation was only observed between MSNs in the same pathway ("homotypic"), and did not affect excitatory transmission onto neurons in the opposite pathway ("heterotypic"). Several observations suggest that this heterosynaptic potentiation is mediated by astrocytes (*8*, *9*). One is that genetic deletion of the calcium receptor inositol-1,4,5-trisphosphate type 2 (IP_3R2), which is required for glial activation, did not prevent homoneuronal DSE, but prevented heterosynaptic potentiation. In addition, activating glia directly by uncaging Ca^{2+} was sufficient to induce heterosynaptic potentiation.

Is circuit-specific potentiation mediated by a single population of astrocytes with subcellular domains targeted to different neuronal subpopulations, or are different populations of astrocytes selectively interacting with D1 or D2 MSNs? To distinguish between these possibilities, Martín *et al.* performed triple whole-cell electrical recordings from two heterotypic MSNs (D1 and D2) and a single astrocyte that was loaded with a Ca^{2+} indicator and a caged Ca^{2+} compound. After identifying whether an astrocyte responded to either the D1 or D2 MSN, a new pair of heterotypic MSNs



Specific astrocytes for specific circuits. Activation of an MSN releases endocannabinoids (eCB), which bind to receptors (CB1) on the presynaptic cortical neuron as well as a functionally connected astrocyte. This decreases neurotransmitter [glutamate (Glu)] release onto the MSN, suppressing its excitation. However, in the astrocyte, CB1 activation triggers a Ca^{2+} signal that causes the release of glutamate. This activates metabotropic glutamate receptors ($mGluR_{1/5}$) on a different cortical neuron and the subsequent release of glutamate onto MSNs in the same population (homotypic), thereby triggering its excitation. This does not affect transmission onto MSNs in a different population (heterotypic).

ILLUSTRATION: C. BICKEL/*SCIENCE*

[1]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA. E-mail: agittis@cmu.edu; dbrasier@cmu.edu

Published by AAAS

was targeted for recordings and the same astrocyte was reactivated by uncaging Ca^{2+}. In every instance, presynaptic glutamate release onto an MSN was potentiated onto the homotypic MSN but not the heterotypic MSN. This firmly established the existence of two distinct subpopulations of astrocytes in the striatum that communicate selectively with distinct populations of MSNs.

Among the questions raised by the findings of Martín *et al.* is the spatial scale over which specific astrocyte-neuron networks operate. Astrocytes can excite one another to propagate signals broadly, but can also operate with synapse-level precision (*10*). Widespread coordination of neurons in the direct or indirect pathways could influence the tone of basal ganglia output under different behavioral conditions, or could contribute to imbalances between these pathways that arise in a number of diseases (*11*). By contrast, local control over small clusters of MSNs would be more likely to influence specific behaviors or drive learning of specific motor skills. Focal stimulation within the striatum can produce movements restricted to certain parts of the body (*12*). Tic disorders have been hypothesized to emerge when small clusters of MSNs, particularly in the direct pathway, become erroneously activated (*13*).

Beyond the striatum, the study by Martín *et al.* raises the possibility that cell-specific astrocyte-neuron networks regulate information flow in many brain areas. Neuronal diversity is essential for creating functionally diverse circuits throughout the brain (*14*, *15*). Although the unique properties and sensitivities of neural circuits have generally been attributed to the properties of their respective neurons, Martín *et al.* raise the intriguing possibility that distinct circuits have dedicated populations of astrocytes acting to regulate their activity, providing a new perspective into the organizing principles of circuit assembly and dynamics throughout the brain. ■

REFERENCES

1. R. Martín *et al.*, *Science* **349**, 730 (2015).
2. A. Volterra, J. Meldolesi, *Nat. Rev. Neurosci.* **6**, 626 (2005).
3. K. Matthias *et al.*, *J. Neurosci.* **23**, 1750 (2003).
4. C. R. Gerfen, D. J. Surmeier, *Annu. Rev. Neurosci.* **34**, 441 (2011).
5. H. H. Yin *et al.*, *Nat. Neurosci.* **12**, 333 (2009).
6. A. C. Kreitzer, R. C. Malenka, *Nature* **445**, 643 (2007).
7. M. Uchigashima *et al.*, *J. Neurosci.* **27**, 3663 (2007).
8. M. Navarrete, A. Araque, *Neuron* **57**, 883 (2008).
9. M. Navarrete, A. Araque, *Neuron* **68**, 113 (2010).
10. A. Araque *et al.*, *Neuron* **81**, 728 (2014).
11. A. C. Kreitzer, R. C. Malenka, *Neuron* **60**, 543 (2008).
12. L. Tremblay, Y. Worbe, S. Thobois, V. Sgambato-Faure, J. Féger, *Mov. Dis.* **30**, 1155 (2015).
13. J. W. Mink, *Arch. Neurol.* **60**, 1365 (2003).
14. R. H. Masland, *Curr. Biol.* **14**, R497 (2004).
15. S. B. Nelson, K. Sugino, C. M. Hempel, *Trends Neurosci.* **29**, 339 (2006).

10.1126/science.aad0678

CLIMATE CHANGE

Has there been a hiatus?

Internal climate variability masks climate-warming trends

By **Kevin E. Trenberth**

Every decade since the 1960s has been warmer than the one before, with 2000 to 2009 by far the warmest decade on record (see the figure). However, the role of human-induced climate change has been discounted by some, owing to a markedly reduced increase in global mean surface temperature (GMST) from 1998 through 2013, known as the hiatus (*1*–*3*). The upward trend has resumed in 2014, now the warmest year on record, with 2015 temperatures on course for another record-hot year. Although Earth's climate is undoubtedly warming, weather-related and internal natural climate variability can temporarily overwhelm global warming in any given year or even decade, especially locally.

> ***"Natural fluctuations are big enough to overwhelm the steady background warming at any point in time."***

Karl *et al.* recently argued that there has been no slowdown in the rise of GMST and hence no hiatus (*3*). The authors compared slightly revised and improved GMST estimates after 2000 with the 1950–1999 period, concluding that there was hardly any change in the rate of increase. Their start date of 1950 is problematic, however. An earlier hiatus, which some now call the big hiatus, lasted from about 1943 to 1975 (see the figure); including the 1950–1975 period thus artificially lowers the rate of increase for the 1950–1999 comparison interval. The perception of whether or not there was a hiatus depends on how the temperature record is partitioned.

Another reason to think there had been a hiatus in the rise of GMST comes from comparing model expectations and observations. Human activities are causing increases in heat-trapping greenhouse gases, mainly carbon dioxide from burning fossil fuels (*4*). These increases are expected to cause rising atmospheric temperatures. Atmospheric aerosols, mostly from fossil fuel combustion, are expected to reduce this rise to some extent. The increasing gap between model expectations and observed temperatures provides further grounds for concluding that there has been a hiatus.

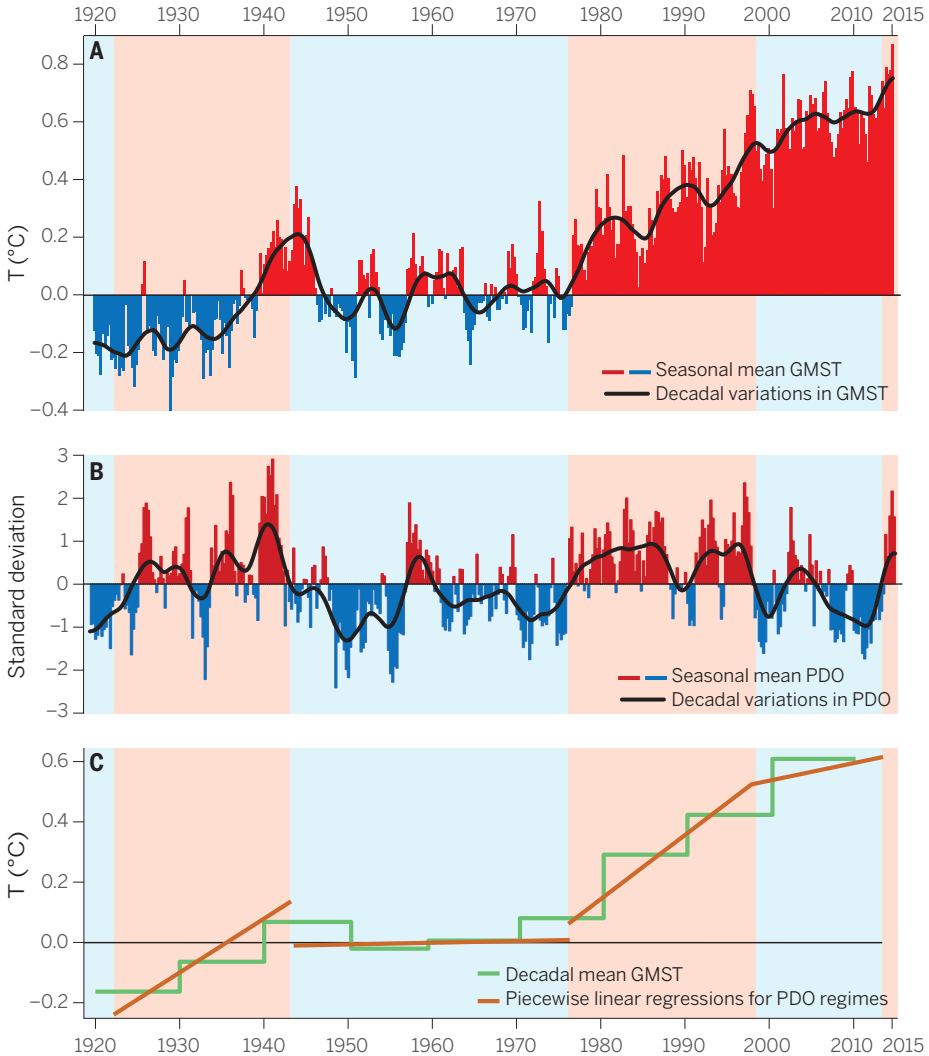
GMST varies from year to year (see the figure) and from decade to decade, largely as a result of internal natural variability. Temperatures have mostly increased since about 1920 and the recent rate is not out of step with the 1950–1999 rate (*3*), but there are two intervals with much lower rates of increase. Only the most recent of these two hiatuses has occurred in the presence of fast-increasing greenhouse gas concentrations. It is thus important to understand its origins and whether or not it indeed indicates a flaw in model projections and thus in climate change theory.

Interannual variability in GMST is partly driven by the El Niño–Southern Oscillation in the Pacific Ocean. The year 1998 was the warmest on rec-ord in the 20th century because of the 1997–1998 El Niño, the biggest such event on record. During that El Niño, ocean heat that had previously built up in the tropical western Pacific spread across the Pacific and into the atmosphere, invigorating storms and warming the surface, especially through latent heat release, while the ocean cooled from evaporative cooling (*5*, *6*). Now, in 2015, another El Niño is under way; it began in 2014 and is in no small part responsible for the recent warmth.

There is also strong decadal variability in the Pacific Ocean, part of which is the Pacific Decadal Oscillation (PDO) (see the figure, panel B). The PDO is closely related to the Interdecadal Pacific Oscillation (IPO) but has more of a Northern Hemisphere focus. Observations and models show that the PDO is a key player in the two recent hiatus periods (*2*). Major changes in tradewinds, sea-level pressure, sea level, rainfall, and storm locations throughout the Pacific and Pacific-rim countries extend into the southern oceans and across the Arctic into the Atlantic (*7*–*9*). The wind changes alter ocean currents, ocean convection, and overturning, for example affecting the Atlantic Meridional Overturning Circulation (*10*). As a result, more heat is sequestered in the deep ocean during the negative phase of the PDO (*1*, *6*, *9*, *11*, *12*). GMST therefore increases during the positive

National Center for Atmospheric Research, Post Office Box 3000, Boulder, CO 80307, USA. E-mail: trenbert@ucar.edu

Published by AAAS



A staircase of rising temperatures. (**A**) Seasonal (December-January-February; etc.) global mean surface temperatures since 1920 (relative to the 20th-century mean) vary considerably on interannual and decadal time scales. Data from (*19*). (**B**) Seasonal mean PDO anomalies (*8*) show decadal regimes (positive in pink; negative in blue) as well as short-term variability. A 20-term Gaussian filter is used in both to show decadal variations, with anomalies reflected about the end point of March to May 2015 (heavy black curves). (**C**) Decadal average anomalies (starting 1921 to 1930) of GMST (green) along with piecewise slopes of GMST for the phases of the PDO (orange). Note how the rise in GMST (A) coincides with the positive (pink) phase (B) of the PDO at the rate given in (C).

phase of the PDO but stagnates during its negative phase (see the figure) (*13*).

Decadal variability also occurs in the Atlantic (*10*, *13*), but the Pacific has dominated recent variability (*1*, *2*, *8*, *9*, *14*, *15*). The Arctic has also seen large changes in recent years, somewhat out of step with the hiatus. However, this region seems to mainly respond to influences from elsewhere, especially the Pacific (*8*, *16*), with snow-ice-albedo feedbacks helping to amplify the changes in surface temperatures (*17*).

There has been considerable speculation about the role of influences external to the climate system on the hiatus. From 1945 to 1970 (*2*, *14*), increases in tropospheric and stratospheric aerosols likely reduced the solar insolation sufficiently to slow warming from increased greenhouse gases. The Clean Air acts of the 1970s in developed countries brought that era to an end. Major volcanic eruptions, especially from Mount Agung (1963), El Chichón (1982), and Mount Pinatubo (1991), had pronounced short-term cooling effects and lowered ocean heat content (*5*). Several small volcanic eruptions (*18*) may have played a role in the 2000s but were not included in IPCC model studies (*6*, *18*). Solar irradiance was slightly lower during the last sunspot minimum (2003 to 2009), and decreased water vapor in the stratosphere after 2000 may have also contributed to decadal variations, but these effects likely accounted for only up to 20% of the recent slowing of the GMST rise (*6*).

Because of global warming, numerous studies have found large regional trends over the past 40 years or so, the period for which we have the best data. However, the associated changes in the atmospheric circulation are mostly not from anthropogenic climate change but rather reflect large natural variability on decadal time scales. The latter has limited predictability and may be underrepresented in many models, but needs to be recognized in adaptation planning. Natural fluctuations are big enough to overwhelm the steady background warming at any point in time.

The main pacemaker of variability in rates of GMST increase appears to be the PDO, with aerosols likely playing a role in the earlier big hiatus. There is speculation whether the latest El Niño event and a strong switch in the sign of the PDO since early 2014 (see the figure) mean that the GMST is stepping up again. The combination of decadal variability and a trend from increasing greenhouse gases makes the GMST record more like a rising staircase than a monotonic rise. As greenhouse gas concentrations rise further, a negative decadal trend in GMST becomes less likely (*13*). But there will be fluctuations in rates of warming and big regional variations associated with natural variability. It is important to expect these and plan for them. ■

REFERENCES AND NOTES

1. K. E. Trenberth, J. T. Fasullo, *Earth's Future* **1**, 19 (2013).
2. A. Clement, P. DiNezio, *Science* **343**, 976 (2014).
3. T. R. Karl *et al.*, *Science* **348**, 1469 (2015).
4. IPCC, *Climate Change 2013: The Physical Science Basis*, T. F. Stocker *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2013).
5. M. A. Balmaseda, K. E. Trenberth, E. Källén, *Geophys. Res. Lett.* **40**, 1754 (2013).
6. K. E. Trenberth, J. T. Fasullo, M. Balmaseda, *J. Clim.* **27**, 3129 (2014).
7. Y. Kosaka, S.-P. Xie, *Nature* **501**, 403 (2013).
8. K. E. Trenberth, J. T. Fasullo, G. Branstator, A. S. Phillips, *Nat. Clim. Change* **4**, 911 (2014).
9. M. H. England *et al.*, *Nat. Clim. Change* **4**, 222 (2014).
10. M. A. Srokosz, H. L. Bryden, *Science* **348**, 1255575 (2015).
11. X. Chen, K. K. Tung, *Science* **345**, 897 (2014).
12. V. Nieves, J. K. Willis, W. C. Patzert, *Science* **349**, 532 (2015).
13. G. A. Meehl, A. Hu, J. M. Arblaster, J. Fasullo, K. E. Trenberth, *J. Clim.* **26**, 7298 (2013).
14. B. A. Steinman, M. E. Mann, S. K. Miller, *Science* **347**, 988 (2015).
15. S. McGregor *et al.*, *Nat. Clim. Change* **4**, 888 (2014).
16. D. Hartmann, *Geophys. Res. Lett.* **42**, 1894 (2015).
17. J. Cohen *et al.*, *Nat. Geosci.* **7**, 627 (2014).
18. B. D. Santer *et al.*, *Nat. Geosci.* **7**, 185 (2014).
19. Data from www.ncdc.noaa.gov/cag/time-series/global, downloaded on 30 June 2015.

ACKNOWLEDGMENTS

Thanks to J. Fasullo and A. Phillips for help with the figure. This work is partially sponsored by U.S. Department of Energy grant DE-SC0012711. The National Center for Atmospheric Research is sponsored by the NSF.

10.1126/science.aac9225

Published by AAAS

VACCINES

Ebola virus vaccines–preparing for the unexpected

Many lives might have been saved if clinical studies of Ebola virus vaccines had been done earlier

By **Hans-Dieter Klenk** *and* **Stephan Becker**

The still ongoing Ebola outbreak in West Africa, which began in 2013, and with more than 27,000 cases and 11,000 deaths so far, highlights the need for a vaccine against the disease (*1*). Hopes to have a vaccine have been nourished in recent years by studies with recombinant vesicular stomatitis virus (VSV) expressing the Ebola virus glycoprotein (VSV-EBOV). VSV-EBOV efficiently protects rodents and nonhuman primates against EBOV from viral strains (Kikwit strain in 1995, for example) that caused past outbreaks, but it was not known if it is also efficacious against the Makona strain responsible for the West African outbreak. On page 739 of this issue, Marzi *et al.* (*2*) demonstrate that the recombinant vaccine provides protective immunity in macaques against the Makona strain. Complete protection was achieved within 7 days after vaccination, suggesting that the vaccine will provide an ideal countermeasure for protecting health care workers and other persons at risk in an outbreak situation.

Analysis by Marzi *et al.* of the mechanisms leading to protection indicates that initially there is a robust innate immune response that slows down virus replication, thereby allowing the generation of antibodies that are the main mediators of protection. The VSV-EBOV vaccine used by Marzi *et al.* is a replicating virus that grows to relatively high titers in vaccinated animals, which has raised questions about its safety. To address this concern, further-attenuated second-generation variants have been developed that cause reduced vaccine-associated viremia, yet also provided complete, single-dose protection of macaques from lethal infection with the Makona strain (*3*). The high potential of the first-generation as well as the second-generation vaccine is underlined by the observation that they protect from infection with Ebola virus variants that express high amounts of a soluble form of the glycoprotein believed to be associated with high viral virulence (*4*, *5*).

Development of vaccines against Ebola virus and its older cousin Marburg virus has a long history, starting shortly after the discovery of these viruses (in 1976 and 1967, respectively). For many years, however, these efforts lacked momentum due to the rare appearance of the outbreaks, which usually were confined to remote areas and involved only a limited number of cases. Furthermore, because there seemed to be no market for the vaccine, support by the pharmaceutical industry for vaccine development was lacking. The situation changed about two decades ago, when bioterrorism was perceived as a realistic threat–Ebola virus ranked at the top of the list of select agents. Of the many different vaccines developed since (*6*, *7*), three have been found in recent months to be safe for human application. They include the VSV-EBOV vaccine and two vaccines based on DNA-adenovirus-vaccinia virus platforms. The thought that many lives might have been saved if the phase 1 clinical studies in which the safety and immunogenicity of the vaccines against Ebola virus was assessed (*8*–*11*) had been performed before and not at the end of the Ebola outbreak is embarrassing (see the figure). Moreover, had the vaccines been available in the early phase of the recent outbreak, it might have been possible to obtain direct information on their impact on survival rates and, thus, on their prophylactic and therapeutic potential. Now, with very few Ebola patients around (the outbreak already peaked, but there are still a few cases arising occasionally), proof of efficacy in humans is difficult to obtain. Knowledge on the efficacy currently depends on data obtained in animal, primarily nonhuman primate,

Institut für Virologie, Philipps-Universität Marburg, Germany.
E-mail: klenk@staff.uni-marburg.de

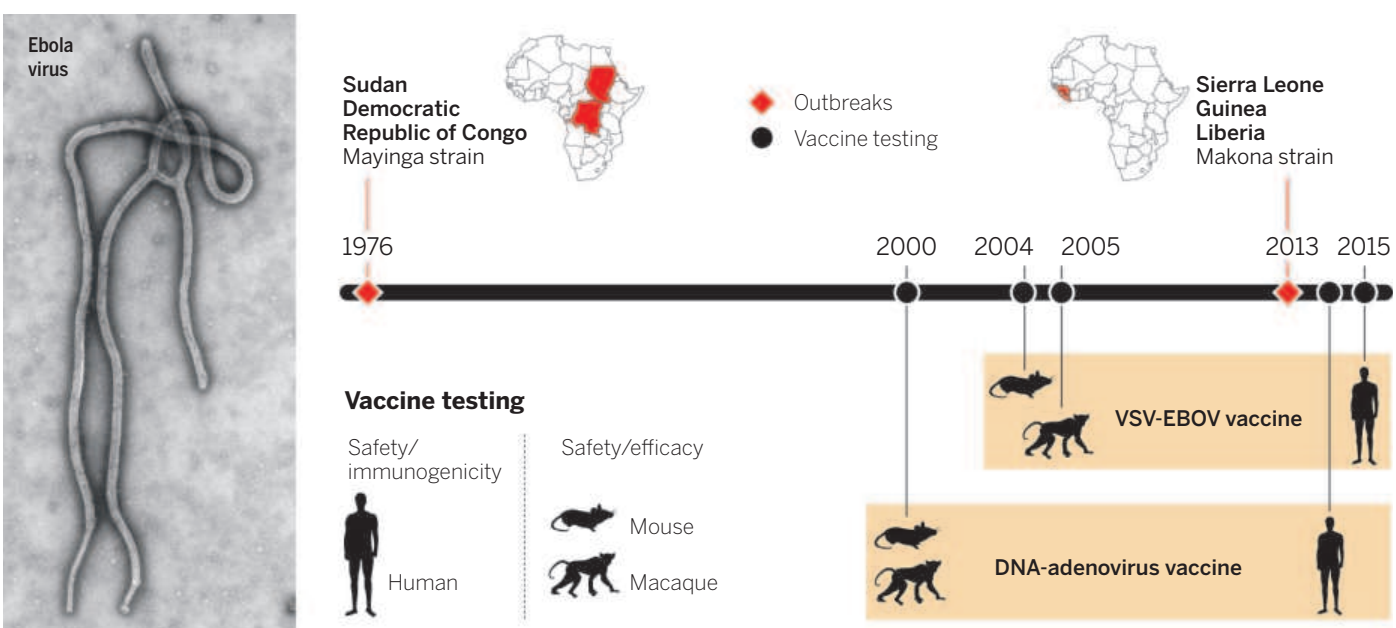


PHOTO: CDC–NAT. CENTER FOR INFECTIOUS DISEASES/SPECIAL PATHOGENS BRANCH; ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Timeline of Ebola virus vaccines. For more than 10 years, it has been known that the adenovirus-based (*12*) and the VSV-based (*13*) Ebola virus vaccines are safe and efficacious in rodents and nonhuman primates, but human safety has been assessed only during the most devastating outbreak in 2013 (*8*–*11*), almost four decades after the first outbreak in 1976.

Published by AAAS

models, although very recent data suggest that VSV-EBOV is efficacious in humans (*14*).

Among the lessons to be learned from the West Africa Ebola outbreak and the history of Ebola virus vaccine development is the need for vaccines (and antiviral treatments) available against this virus and other reemerging viruses, such as Marburg virus, severe acute respiratory syndrome (SARS) coronavirus, Middle East respiratory syndrome (MERS) coronavirus, Crimean-Congo hemorrhagic fever virus, and Rift Valley fever virus, at the onset of an outbreak in sufficient quantities to allow immunization of all persons with a high risk to be exposed to infection. To this end, the vaccines must be developed at least to a stage where safety for human application has been proven (phase 1 clinical trials). Facing the difficulties to show efficacy of those emergency vaccines, correlates for human protection should be sought to assess vaccine efficacy under pre-outbreak conditions, and platforms must be developed that allow the fast generation of vaccines against a so far unknown emerging virus. The development and production of such vaccines require public funding. ■

"Among the lessons to be learned from the West Africa Ebola outbreak ... is the need for vaccines (and antiviral treatments) available against this virus and other reemerging viruses ... at the onset of an outbreak ..."

REFERENCES

1. H. D. Klenk, *Nat. Med.* **6**, 1322 (2000).
2. A. Marzi *et al.*, *Science* **349**, 739 (2015).
3. C. E. Mire *et al.*, *Nature* **520**, 688 (2015).
4. G. S. Mohan, W. Li, L. Ye, R. W. Compans, C. Yang, *PLOS Pathog.* **8**, e1003065 (2012).
5. V. A. Volchkova, O. Dolnik, M. J. Martinez, O. Reynard, V. E. Volchkov, *J. Infect. Dis.* **204** (suppl. 3), S941 (2011).
6. A. Marzi *et al.*, *Science* **348**, 439 (2015).
7. C. E. Mire, T. W. Geisbert, in *Biology and Pathogenesis of Rhaddo- and Filoviruses*, A.K. Pattnaik, M. A. W. Whitt, Eds. (World Scientific, Singapore, 2015), pp. 587–620.
8. S. T. Agnandji *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1502924 (2015).
9. J. E. Ledgerwood *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1410863 (2015).
10. T. Rampling *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1411627 (2015).
11. J. A. Regules *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1414216 (2015).
12. N. J. Sullivan, A. Sanchez, P. E. Rollin, Z. Y. Yang, G. J. Nabel, *Nature* **408**, 605 (2000).
13. S. M. Jones *et al.*, *Nat. Med.* **11**, 786 (2005).
14. A. M. Henao-Restrepo *et al.*, *Lancet* 10.1016/S0140-6736(15)61117-5 (2015).

10.1126/science.aad0681

CANCER

Tumor angiogenesis, from foe to friend

Treatments that normalize or even promote blood vessel growth may enhance drug delivery to tumors

By **Lee B. Rivera** *and* **Gabriele Bergers**

Targeting the tumor vasculature to "starve a tumor to death" instead of targeting tumor cells with chemotherapeutic drugs was conceived over four decades ago and has led to the development of antiangiogenic drugs approved for use against various human malignancies (*1*). So far, however, antiangiogenic therapy has not fulfilled expectations because it aids only a subset of cancer patients and provides only transitory improvements. Vascular-disrupting agents were developed to more rigorously deplete tumor vessels (*2*). However, this approach leads to hypoxia, which promotes neovascularization and tumor regrowth. The sobering realization is that the more we try to exterminate tumor vessels, the more aggressively tumors respond to impede these efforts, sometimes becoming more belligerent tumors. Is manipulating the vasculature to control tumor growth a promising strategy after all?

Conventional angiogenic inhibitors primarily block the vascular endothelial growth factor (VEGF) signaling pathway, and prune rather than eradicate tumor vessels. By contrast, some flavonoids and microtubule-binding molecules rapidly and selectively kill the tumor vasculature (*2*). The latter, wide-scale vessel destruction results in a more massive necrosis and hypoxia than occur with antiangiogenic therapy, and therefore is effective at killing the bulk of the tumor. This reaction seems to be primarily driven by low oxygen tension. Conversely, hypoxia also triggers several resistance mechanisms that drive tumor regrowth. It promotes the epithelial-mesenchymal transition and stemlike properties of tumor cells, increases the expression of proangiogenic and invasive factors, and drives the infiltration and polarization of angiogenic and immune-suppressive myeloid cells (*1*, *3*).

The approach of eradicating tumor vessels therefore seemed to be at a dead end until a new opportunity arose with the concept of vessel normalization. Detailed analysis of the changes resulting from inhibition of VEGF and the VEGF receptor (VEGFR) revealed that antiangiogenic therapies leave behind a more mature and functional vasculature by selectively pruning immature blood vessels (*4*). This results in enhanced oxygenation and perfusion throughout the tumor and subsequent activation of antitumor immunity, thus creating conditions for better drug delivery and efficacy. Vascular normalization with VEGF and VEGFR inhibitors has indeed been demonstrated in several preclinical cancer models. However, the inability to finely tune antiangiogenic therapy to create persistent normalization without further pruning leads to a recurrence of hypoxia and the emergence of acquired resistance.

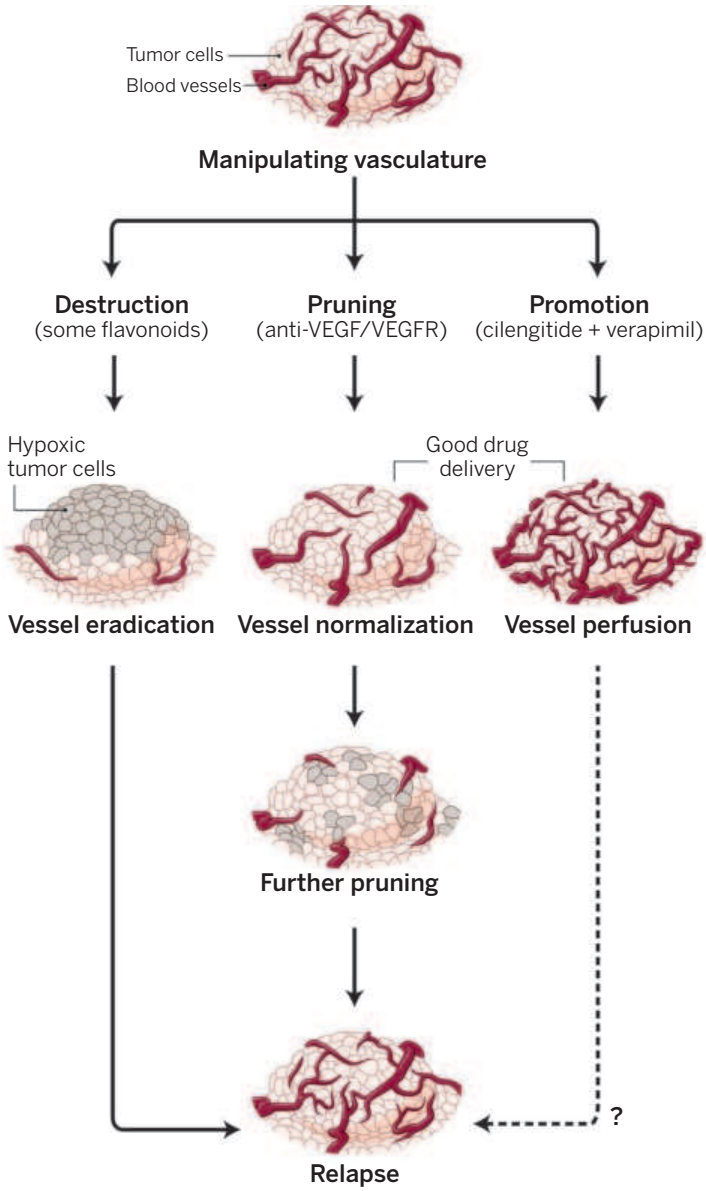
Recently, use of an alternative strategy showed that the antimalarial agent chloroquine decreased tumor blood vessel tortuosity and vessel density while increasing endothelial cell organization, the coverage of endothelial cells by pericytes, and vessel perfusion (*5*). This enhanced tumor oxygenation and increased efficacy of chemotherapy. Chloroquine's alkalinizing nature hinders the endosomal cycling process, a mechanism that controls the trafficking of surface proteins between the cytoplasm and cell surface membrane. Indeed, the compound's capacity to normalize the vasculature was due to abnormal cycling and signaling of the surface protein Notch1, a negative-regulator of vascular sprouts. Vessel normalization to enhance vascular function has thus become an important concept, with the goal of improving delivery of chemotherapy and promoting a more oxygenized and immune-stimulating tumor environment.

The idea of manipulating the tumor vasculature recently advanced a step further with the demonstration that stimulating tumor angiogenesis can sensitize tumors to chemotherapy (*6*). The study took advantage of the angiogenic properties of cilengitide, a cyclic pentapeptide that binds to the integrins αVβ3 and αVβ5 and blocks cells adhesion. It also used verapimil, a Ca^{2+}-channel blocker. Cilengitide, when administered at low doses, can alter VEGF receptor trafficking

Department of Neurological Surgery, Brain Tumor Research Center, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA.
E-mail: lee.rivera@ucsf.edu; gabriele.bergers@ucsf.edu

Published by AAAS

on endothelial cells to favor VEGF-induced angiogenesis (*7*), and verapimil induces vasodilation and subsequent blood flow. The combination of verapimil and cilengitide increased vessel density, dilation, permeability, and perfusion within tumors, which in turn increased tumor oxygenation and enabled more efficient delivery and enhanced efficacy of the chemotherapeutic drug gemcitabine. Importantly, the beneficial effects in rodent models of cancer were not just obtained by better drug delivery to the tumor site but also by a more efficient drug influx into the tumor cells and its subsequent activation. Gemcitabine uptake into cells is regulated by equilibrative nucleoside transporters 1 and 2 (ENT1 and ENT2) and by concentrative nucleoside transporter 3 (CNT3), which mediates the unidirectional flow of the drug into the cell. On entering the tumor cell, gemcitabine is converted into its active form by deoxycytidine kinase (DCK). Treatment with the triple combination of verapimil, cilengitide, and gemcitabine enhanced ENT1, ENT2, CNT3, and DCK expression, in part by lowering hypoxia and thereby increasing the intratumoral influx of gemcitabine (*6*).



Exploiting the vasculature to control tumors. Current antiangiogenic approaches aim to sustain vessel normalization—the first phase of blood vessel pruning (before the onset of increased hypoxia)—for enhanced oxygenation, drug delivery, and efficacy. These goals are shared by vascular promotion therapy. So far, antiangiogenic and vessel-disrupting agents have led to tumor relapses caused by reneovascularization or altered tumor behavior. It remains to be seen whether vascular promotion follows a similar path.

The success of the vascular promotion approach both in tumor models that were either already angiogenic (Lewis lung cancer) or hypovascular (a genetic model of pancreatic adenocarcinoma that responds poorly to its standard-of-care gemcitabine) suggests a more general applicability. A better mechanistic understanding of this strategy is needed to move vascular promotion forward to the clinic. Clearly, gemcitabine appears to be the drug of choice at this stage due to its enhanced uptake and activation, whereas efficacy of another chemotherapeutic drug, cisplatin, was only improved as a result of better drug delivery to the tumor.

Can other chemotherapeutic drugs that enter the cells via ENTs and CNTs be used with this approach? Are other drug transporters regulated by hypoxia and if so, which chemotherapeutics would then most benefit from this strategy? Given that celingitide can have both pro- and antiangiogenic effects in a concentration-dependent manner that may not be easily assessed in the clinic, alternative approaches to promote angiogenesis in a controlled manner should be considered. Also unclear is whether enhanced angiogenesis is even needed in already highly angiogenic tumors or if verapimil would suffice to improve blood flow and oxygenation. Regardless, the idea of vascular promotion supports the important revelation that one can target the vasculature but leave it intact and thereby provide therapeutic benefits. This has been successfully exploited to reveal angiocrine signaling cues between endothelial cells and tumor cells that promote tumor growth and invasion (*8*). For example, lymphoma-derived fibroblast growth factor 4 induced expression of the Notch ligand Jagged 1 in endothelial cells, which in turn induced Notch-dependent lymphoma invasion and resistance to chemotherapy; deletion of the gene encoding focal adhesion kinase in endothelial cells led to a reduction in several secreted factors without affecting vessel density and enhanced tumor sensitivity to chemotherapies (*9*, *10*). These results suggest that targeting angiocrine factors can provide therapeutic benefits without affecting vascular function and potentially be exploited for vascular promotion therapy.

Concerns about the nature of tumor relapse from vascular promotion should be investigated, specifically if influx of chemotherapeutic drugs into tumor cells stops due to other mechanisms. This could theoretically lead to a fast regrowth of the tumor with potentially enhanced metastasis as a likely by-product of vessel leakiness and tumor angiogenesis. Designing trials of vascular promotion for tumors for which gemcitabine has been standard care would be a first important step toward clinical validation. If successful, vascular promotion in such or other combinations—for example, with immune therapies—could prove to be an important new interventional approach in cancer therapy. ■

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

REFERENCES

1. R. K. Jain, *Cancer Cell* **26**, 605 (2014).
2. A. Hollebecque, C. Massard, J. C. Soria, *Curr. Opin. Oncol.* **24**, 305 (2012).
3. G. Bergers, D. Hanahan, *Nat. Rev. Cancer* **8**, 592 (2008).
4. R. K. Jain, *Nat. Med.* **7**, 987 (2001).
5. H. Maes *et al.*, *Cancer Cell* **26**, 190 (2014).
6. P. P. Wong *et al.*, *Cancer Cell* **27**, 123 (2015).
7. A. R. Reynolds *et al.*, *Nat. Med.* **15**, 392 (2009).
8. J. M. Butler, H. Kobayashi, S. Rafii, *Nat. Rev. Cancer* **10**, 138 (2010).
9. Z. Cao *et al.*, *Cancer Cell* **25**, 350 (2014).
10. B. Tavora *et al.*, *Nature* **514**, 112 (2014).

10.1126/science.aad0862

Published by AAAS

QUANTUM OPTICS

The on-ramp to the all-optical quantum information processing highway

An optical waveguide circuit can be flexibly programmed with near-perfect fidelity

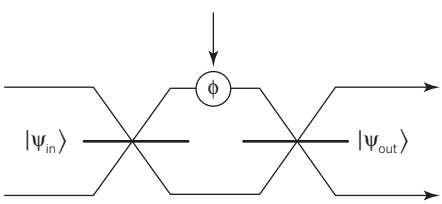
By **Peter P. Rohde**[1]
and **Jonathan P. Dowling**[2]

Since the first formulations of quantum mechanics in the early 20th century, it became clear that the enormous complexity of quantum-mechanical systems presented intractable computational problems. Richard Feynman was the first to turn this problem on its head. He asked whether we could exploit this quantum complexity to construct a computer based on these same quantum mechanical principles, offering exponential algorithmic improvements, and whether such a computer could efficiently simulate quantum systems that our classical computers are unable to simulate. This challenge initiated the field of quantum computing and is today a major field of research in the physics and computer science communities. One hurdle has been to construct devices that match the flexible programmability of classical computers. On page 711 of this issue, Carolan *et al.* (*1*) present a step in that direction, a fully reconfigurable optical waveguide circuit that can be programmed to implement arbitrary linear optics transformations on up to six optical modes.

A first quantum algorithm, proposed by Deutsch and Jozsa (*2*), showed that there is in fact an exponential separation in computational power between classical and quantum computers. This finding invigorated the field and stimulated pioneering research into other quantum algorithms, of which there are now many. In principle, any well-controlled quantum system could be used as a basis for constructing a quantum computer. In practice, optical quantum computing (*3*) has emerged as a leading candidate because of the relative ease of preparing, manipulating, and detecting photons, as well as their long decoherence times.

Elementary demonstrations of optical quantum gates have been demonstrated many times (*4*). The challenge now is to scale up this technology to enable interesting computational algorithms that can outperform our best classical computers. Not only will such scalable devices need to be able to process photons by the thousands but also the network of interference between them will need to be reconfigurable, such that they can be programmed to implement the full plethora of useful quantum algorithms.

Carolan *et al.* demonstrate reconfigurability of a number of elementary quantum information processing protocols: An entangling quantum gate (a building block for constructing larger quantum circuits); the preparation of a small so-called "cluster state" (a type of state that may be used to implement quantum algorithms); quantum process tomography (a procedure by which the operation of a quantum device can be characterized); and boson sampling (a simple but nonuniversal model for optical quantum computing). Their device was reconfigured with more than 100 settings, demonstrating its versatility, with near-perfect fidelities across the board.



The basic quantum building block. Carolan *et al.* assembled optical waveguide circuits out of switching components that take a quantum state ψ_{in} to ψ_{out} via a phase shift of ϕ.

Although a six-mode optical system is certainly far from being able to outperform existing classical computers, this experiment provides a proof-of-principle that miniaturized and arbitrarily reconfigurable devices may be constructed, with very high fidelities, for the purposes of optical quantum information processing tasks. However, many engineering challenges remain for the future development of the field. For the purposes of "universal" quantum computing, such a device must dynamically reconfigure itself based on partial measurements within the duration of the computation. This capability is known as "fast-feed-forward," in which the device must be able to measure a subset of the photons and reconfigure the remainder of the circuit depending on the measurement outcomes. Switching times on the order of nanoseconds will be needed, because those in the experiment of Carolan *et al.* were on the order of seconds.

Nonetheless, even in the absence of fast-feed-forward, reconfigurable waveguides still have some very important uses today. The most prominent recent example of this is "boson-sampling" (*5*), a simplified model for optical quantum computing in which no feedforward is required. Instead, photons propagate through a fixed, passive network and are subsequently measured. Although this protocol is not believed to be universal for quantum computing and can implement only a single very specific task, it is believed to implement a computational problem intractable on classical computers. One hundred independent instances of this problem were demonstrated by Carolan *et al.* by reconfiguring the same device.

A second use for passive networks is a recent quantum metrology technique (*6*). It outperforms all classical metrological approaches with only single photons and a fixed network of a particular structure. The device of Carolan *et al.* could easily be programmed to implement this method.

Scalable optical quantum information processing remains an extremely formidable long-term goal. Developing the basic technological building blocks is therefore a pressing goal. The work by Carolan *et al.* demonstrates the viability of one of these essential technological requirements, reconfigurable optical circuits, which, in the long term and with other technological advances, will facilitate larger scale and more versatile quantum information processing protocols but even in the short-term will find several important uses. ■

[1]*Centre for Quantum Computation and Intelligent Systems (QCIS), Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia.*
[2]*Hearne Institute for Theoretical Physics and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA.*
E-mail: dr.rohde@gmail.com; jdowling@lsu.edu

REFERENCES

1. J. Carolan *et al.*, *Science* **349**, 711 (2015).
2. D. Deutsch, R. Jozsa, *Proc. R. Soc. Lond. A* **439**, 553 (1992).
3. E. Knill, R. Laflamme, G. J. Milburn, *Nature* **409**, 46 (2001).
4. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, D. Branning, *Nature* **426**, 264 (2003).
5. S. Aaronson, A. Arkhipov, *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, 2011), pp. 333–342.
6. K. R. Motes *et al.*, *Phys. Rev. Lett.* **114**, 170802 (2015).

10.1126/science.aac7250

Published by AAAS

ECOLOGY

A year of change

A naturalist turns a reverent eye to the birds of the Scottish highlands

By **Christopher Kemp**

Every day from my work desk, I look upon a boring and uneventful tableau: a cluster of gray and brown buildings, some of them tall; half-filled parking lots; and rooftop duct-work. It is almost unremittingly without consequence, but occasionally, a peregrine falcon slices through it like a gray blade.

Books like *Gods of the Morning* by John Lister-Kaye make the waiting between sightings easier. From his remote, altitudinous home in Scotland, where he operates the Aigas Field Centre, the English naturalist has written a magical book—a poetic celebration of the many different birds that own the wild air around him.

The reviewer is in the Department of Translational Science and Molecular Medicine, Michigan State University, Grand Rapids, MI 49503, USA. E-mail: cjkemp@gmail.com

For almost 40 years—since the center was opened in 1977—Lister-Kaye has perched with his family at Aigas, watching the seasons change. For decades, the different bird species have arrived and departed again with such cyclical regularity that Lister-Kaye could tell what month it was by their movements. No longer.

First, he finds a lonely and diminutive blackcap in September. It had flown into a window and died. The rest of the blackcaps had left already, migrating south to Spain and Italy and warmer temperatures. This one had stayed behind. Twenty-five years earlier, blackcaps themselves had heralded a change in the glen, arriving unexpectedly. Now, Lister-Kaye wonders if they have begun to miss their migratory cues to leave when the weather grows cold.

Next, the rooks appeared a full season early in a noisy black tangle to take up their usual nesting spots in the lime trees. Lister-Kaye writes: "What had brought them to my bathroom window, to dance and haggle through the un-leafing tree tops, to soar and plunge and cry among the striping rays of the lowering sun? Just what else was going on?" What follows is Lister-Kaye's year-long rumination on the changes he has seen—and continues to see—and the factors that might be causing them.

As the years have passed, Lister-Kaye has watched the waders gradually disappear from the moorland—along with the curlew, the lapwing, the greenshank, and the others. The short-eared owls are gone from the grassland; the hen harriers are gone from the meadow; the oystercatchers are gone from the river.

Gods of the Morning
A Bird's-Eye View of a Changing World
John Lister-Kaye
Pegasus, 2015. 304 pp.

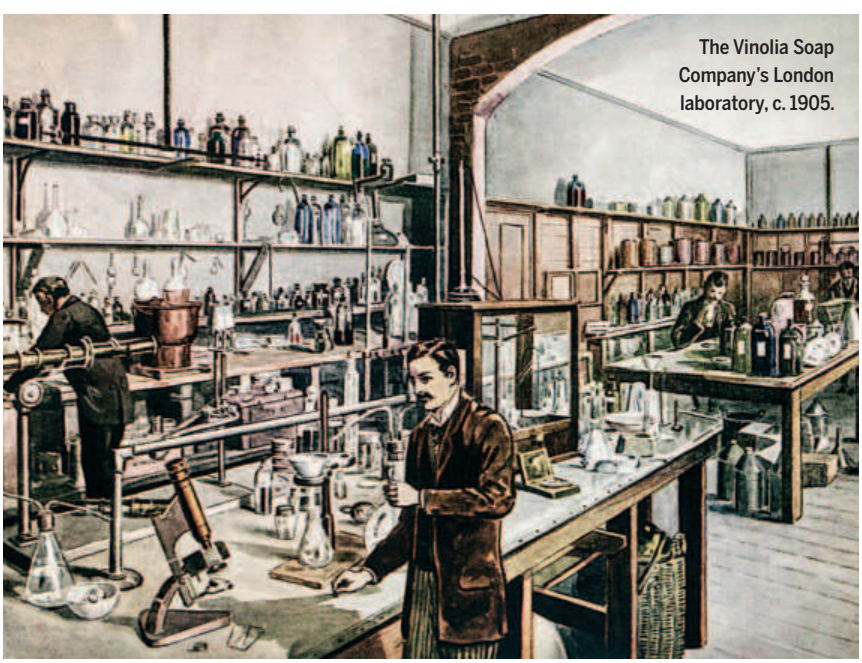
In the prologue, he admits that he had not planned to write this book; instead, he was compelled by the environmental changes he has observed. Wary of blaming what he sees squarely on climate change, Lister-Kaye writes: "But I cannot deny that in the last few years it would appear that the pace of climate change has accelerated and we have entered a period of total weather unpredictability."

Lister-Kaye watches the greylag geese pass overhead in V-shaped formation on their way to Iceland to breed. They are a month late. The swallows arrive from Africa a month early. Can the birds handle the changes?

In weighty, burnished prose, Lister-Kaye describes it all unflinchingly: the change, the loss, the disruption, the bodies. It is an old man's book. Lister-Kaye, who is 69 years old, is a wise and gentle soul, at one with the land he has lived on for 40 years. His wisdom is present on every page.

Gods of the Morning is a lovely ode—both mournful and hopeful—to the snow-choked valleys and the birds that live there, prevailing against the innumerable odds.

10.1126/science.aac8251



The Vinolia Soap Company's London laboratory, c. 1905.

PHOTO: PHOTOS.COM/THINKSTOCK

CHEMISTRY

The Matter Factory
A History of the Chemistry Laboratory
Peter J. T. Morris
Reaktion Books, 2015. 416 pp.

Just say the word "chemistry," and it evokes an image of bubbling flasks, Bunsen burners, and shelves filled with dusty reagents. Yet, few histories of chemistry have paid much attention to the laboratory itself, focusing instead on tales of chemists and their research. Arguing that changes in laboratory design were critical to enabling the progress of chemistry, Peter J. T. Morris explores the origins and evolution of the chemistry laboratory, from medieval alchemy dens through today's state-of-the-art facilities. Rich in detail and featuring an array of engravings, illustrations, and photographs, *The Matter Factory* is an unusual and engaging history.

10.1126/science.aad0929

Published by AAAS

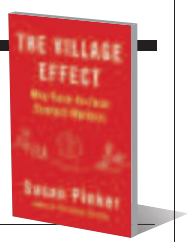
SOCIOLOGY

The original social network

Forging personal connections in the digital age

By **Barry Wellman**

The Village Effect
Why Face-to-Face Contact Matters
Susan Pinker
Atlantic Books, 2015. 432 pp.



In *The Village Effect*, Susan Pinker charges directly into the fierce debate that has raged for centuries over the fate of civilization. Put simply, the debate is about whether relationships are falling apart—at home, in communities, and at work. On the one hand, doomsayers have lamented the supposedly destructive effect of large-scale social changes on warm, supportive relationships. With little or no evidence, they have deduced that such changes—be they bureaucratization, industrialization, technological change, or urbanization—have withered social ties. On the other hand, ethnographers and survey researchers have spent decades amassing evidence that strong and weak ties continue to thrive in communities, families, and workplaces.

Along came the Internet in the early 1990s, followed two decades later by smart phones. Not surprisingly, the widespread use of these media has led to fears that they are sucking us into digitally borne ties that afford only weak simulacra of meaningful relationships. With so much digital activity, does real communion with family, friends, and workmates fade away? By now, there is considerable evidence that digital media do not replace face-to-face relationships, nor do they wither communities. Rather, digital media maintain, complement, and extend relationships when people are not able to connect in person. And even when kith and kin are near at hand, people use digital media to make appointments, share information, and exchange parting comments.

Susan Pinker's entry into the debate adds developmental psychological knowledge to the social scientific mix. In a lively book written for an intelligent general audience, she uses a mixture of human-interest stories and summaries of scholarly evidence to show the importance of in-person ties. She does a commendable job of finding expertise and stories about a variety of subjects, including family relations, neighboring, well-being, social eating, child-rearing, longevity, old age, and contagious ideas and behavior. There are scores of endnotes, most to the scholarly literature.

Pinker's message is complex. First, she uses psychological data to show that people want face-to-face relationships. She presents a variety of evidence to show that the brain is social, evolving at a time when social cohesion meant survival, while social isolation meant starvation and death. Second, she argues that people are happier, healthier, and perform better when they are in face-to-face relationships and that such relationships outperform online connectivity. In one of her many examples, Sandy Pentland and Ben Waber used sociometric badges—wearable electronic devices that measure gestures, vocal features, and physical proximity to others—to identify work groups' cohesion, cleavages, leadership, and productivity. The researchers found, for example, that the more people talk with others on their teams, the more productive they are.

The Village Effect **argues that face-to-face relationships are critical to our health, happiness, and productivity.**

Her third argument is the most daring—and the most troublesome. Pinker argues that we need—and are in part recreating—village-like structures and dynamics of densely knit supportive relationships. She takes her cue from Hillary Rodham Clinton's *It Takes a Village*, which usefully tried to move public dialogue beyond the notion that individual households are each separately responsible for how their children are raised and turn out (*1*).

Yet, Pinker and Clinton take the village focus too far, ignoring the fact that families and communities no longer operate from castle-like homes in all-embracing rural and urban villages. I wish that more attention had been paid to analyzing representative examples, rather than focusing on pastoralist nostalgia. For instance, Pinker spends many pages extolling the supportiveness of an isolated Sardinian village filled with centenarians. Yet, this and similar accounts are not representative examples of how most people live—and judging from the out-migration from there, not how most people want to live.

Families in Western societies are more apt to spend time networking as semiautonomous individuals with separate, but coordinated, agendas. They use a multitude of transportation and communication services to connect with household members, kin, friends, co-workers, and schoolmates. To be sure, face-to-face contact is important, but so are phones, the Internet, cars, and planes—all of which bring people together and maintain relationships when they are physically apart. By being networked rather than village-bound, people gain discretion and flexibility in how they live and work together.

Nevertheless, whether schmoozing in villages or online, Pinker provides important evidence that abundant interpersonal contact can make us "healthier, happier, and smarter." The book reminds us of the pressing need for more high-quality analysis of how face-to-face contact compares with, replaces, and complements the widespread use of digital media. Yet, outside of nostalgia for village life, we do not get much sense of the social transformations accompanying the current great turn to digital media.

The reviewer is the coauthor of Networked: The New Social Operating System *(MIT Press, Cambridge, 2012). E-mail: wellman@chass.utoronto.ca*

REFERENCES

1. H. R. Clinton, *It Takes a Village: And Other Lessons Children Teach Us* (Simon & Schuster, New York, 1996).

10.1126/science.1261448

CREDIT: TUKARAM KARVE/SHUTTERSTOCK

Published by AAAS

LETTERS

Edited by **Jennifer Sills**

Parks for science, science for parks

IN THEIR INSIGHTFUL Editorial "Parks for science" (19 June, p. 1291), G. Machlis and M. McNutt urge scientists to make greater use of protected areas for their research. The benefits are manifold, they argue, and include prospects for exciting new discoveries, the use of parks as experimental controls, and improved park management based on the data collected. There is another advantage that in many cases is even more compelling: Parks that are the foci for research tend to be better protected (*1*).

Especially in developing nations, parks hosting research often suffer less poaching, illegal logging, and illicit mining. Scientists are not just passive but also active park defenders, as shown by the strong avoidance of areas frequented by scientists by poachers and encroachers (*1*–*3*). Furthermore, monies from research tend to be spent locally, benefiting nearby communities and thereby fostering economic support for protected areas (*4*). Scientists are often staunch defenders of parks politically, encouraging journalists to visit and write about the parks, and lobbying both locally and internationally—often with great effectiveness—for their protection (*1*).

Yes, parks offer many research opportunities for scientists. But scientists also offer many potential protective benefits for parks.

William F. Laurance

Centre for Tropical Environmental and Sustainability Science and College of Marine and Tropical Biology, James Cook University, Cairns, QLD 4878, Australia. E-mail: bill.laurance@jcu.edu.au

REFERENCES

1. W. F. Laurance, *Trends Ecol. Evol.* **28**, 261 (2013).
2. G. Campbell *et al.*, *Biol. Lett.* **7**, 723 (2011).
3. N. Tagg *et al.*, *Animal Conserv.* 10.1111/acv.12212 (2015).
4. C. A. Kirkby *et al.*, *PLOS ONE* **5**, e13015 (2010).

Disaster response: Volunteers, unite!

TWO MASSIVE EARTHQUAKES that caused 8898 deaths, 22,309 injuries, and billions of dollars damage in Nepal (*1*) prompted substantial support from aid agencies, relief organizations, and volunteers for rescue and relief. Volunteers in scientific communities—affiliated with government, academic, and private research institutions, as well as individual researchers and citizen scientists around the world—provided free satellite imagery, helped map the damage, and analyzed the disaster data (*2*). Nepal has weak data infrastructure and limited scientific capacity. The efforts of science volunteers provided crucial data for rescuers and disaster responders, informed the public about the scale of the damage, and prevented further damage.

Scientific volunteerism during crises has emerged spontaneously as the Internet, smartphones, and social media have fueled a feeling of shared responsibility. However, working sporadically in cyberspace might lead to duplication of tasks and an overload of information. Poorly equipped government authorities, as in Nepal, cannot easily find useful information in the gigantic pool of data. The services offered

ONLINE BUZZ: CAREERS

Career success: Updating the battle plan

In his Working Life column, "Getting noticed is half the battle" (10 July, p. 206), E. P. Diamandis attributed his successful science career to his early-career decisions, including attending seminars to be noticed by potential employers and working long hours while his wife cared for their children. In the online comments section, readers passionately reacted to his strategies and *Science*'s decision to publish the piece. Excerpts of their responses are below. Read the full comments, and add your own, at http://comments.sciencemag.org/content/10.1126/science.349.6244.206.

A selection of your thoughts:

...It seems that the effort of being noticed by the powerful is the most important driver of the author. Forget about collaborations, sharing with colleagues, helping younger colleagues, discussing ideas. Intervening in lectures with the only purpose of being noticed sounds Machiavellian.... This kind of "suggestion" will only serve to discourage those who are in science for the love of knowledge...to the advantage of those who are...hyper-competitive, self-centered, egoistic (and sexist).... ***Monica Zop***

This is why I left academia. There are many interesting businesses that will hire you to a permanent position after a straightforward 30-minute interview. The wages will be higher than academia, and in my experience you may have more freedom to pursue interesting projects and develop your professional skills.... ***David Blake***

...[H]ere we have Exhibit A for why women disproportionately drop out of the pipeline post-Ph.D. If getting noticed is half the battle when you have the luxury of dumping all your family responsibilities on your wife, exactly what proportion of the battle do you think it might be for *her* career?.... ***Sarah Jordan***

It would seem...that [Diamandis's] wife should have been chosen to write an article on career success, given that she had to take care of all his domestic chores and she still became a senior scientist at a major teaching hospital. ***Anonymous***

If the [spouse] were a man, he would have the opposite problem: societal pressure to...be the breadwinner. Not because he wants to be, but because stereotypical gender roles are still around.... ***Anonymous***

...Diamandis...was merely pointing out that his wife had to handle everything at home so that he could work 16-hour days,...not...that women should always do this....The real point of the article is...the inequities in a system that forces one spouse...to work 16-hour days so that the other spouse is forced to sacrifice his or her career.... ***Warm Neutron***

...The author's advice to work hard, publish good science, and build professional networks is timeless and sound. The remainder of his suggestions (unsustainable work hours, sycophantic behaviors, gendered career expectations) are not things I would advise scientists of any age. ***Katherine H. Freeman***

Published by AAAS

Debris surrounds the temple of Seto Machindranath in Jana Bahal 3 months after the Nepal earthquake.

PHOTO: RATNAYAKE/REX SHUTTERSTOCK/AP IMAGES

by science volunteers could even make a situation worse.

Given that the trends of scientific volunteerism have been increasing globally during disasters, from Typhoon Haiyan in Philippines (*3*) to the recent earthquake in Nepal, a global action to systematize these efforts is necessary. Currently, this increasing global trend of volunteerism in scientific communities during natural disasters is not acknowledged in the Sendai Framework for Disaster Risk Reduction 2015–2030—a global framework for disaster risk reduction (*4*). We need a strategy to harness the efforts of scientific volunteers in future. To be better equipped for the next disaster, we should analyze the patterns of science volunteers, including their affiliations and the nature of their involvement; prepare government authorities, especially those in developing countries, to harness invaluable information provided by volunteer scientists; and create a global network of volunteer scientists that could contribute to a coordinated effort.

Uttam Babu Shrestha,[1]*
Sujata Shrestha,[2] Achyut Aryal[3]

[1]Institute for Agriculture and the Environment (IAgE), University of Southern Queensland, Toowoomba, QLD 4350, Australia. [2]Department of Biology, University of Massachusetts Boston, Boston, MA 02125, USA. [3]Institute of Natural and Mathematical Sciences, Massey University, Albany 0745, Auckland, New Zealand.

*Corresponding author.
E-mail: ubshrestha@yahoo.com

REFERENCES

1. Government of Nepal, Nepal Disaster Risk Reduction Portal (http://drrportal.gov.np).
2. Center for Geographic Analysis, Harvard University, Nepal Earthquake—Geographic Community Response (http://gis.harvard.edu/services/project-consultation/project-resume/nepal-earthquake-geographic-community-response).
3. D. Butler, "Crowdsourcing goes mainstream in typhoon response," *Nature* 10.1038/nature.2013.14186 (2013).
4. United Nations Office for Disaster Risk Reduction, Post-2015 Framework for Disaster Risk Reduction (www.unisdr.org/we/coordinate/hfa-post2015).

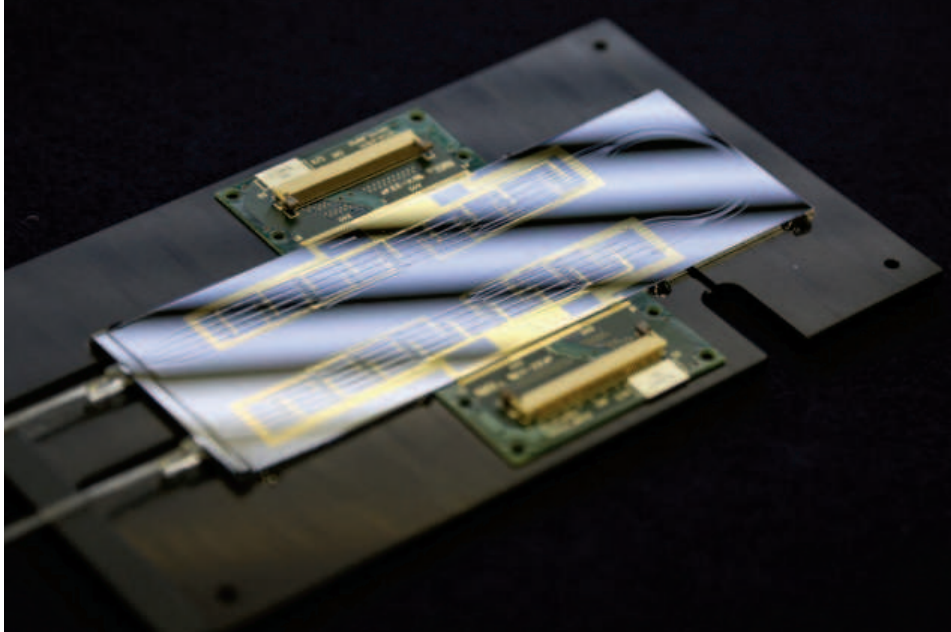
RESEARCH

Glial networks get cell specific

Martín et al., *p. 730*

IN *SCIENCE* JOURNALS

Edited by **Nick Wigginton**



This versatile universal linear optics chip is reprogrammable

QUANTUM OPTICS

Complex quantum optical circuitry

Encoding and manipulating information in the states of single photons provides a potential platform for quantum computing and communication. Carolan *et al.* developed a reconfigurable integrated waveguide device fabricated in a glass chip (see the Perspective by Rohde and Dowling). The device allowed for universal linear optics transformations on six waveguides using 15 integrated Mach-Zehnder interferometers, each of which was individually programmable. Functional performance in a number of applications in optics and quantum optics demonstrates the versatility of the device's reprogrammable architecture. — ISO

Science, this issue p. 711; see also p. 696

EBOLA VACCINE

Shortening the time to protection

Although Ebola vaccine candidates have entered clinical trials in West Africa, there is little information available on the mechanism of protection. A single dose of the recombinant vesicular stomatitis virus–Ebola vaccine protects nonhuman primates, acting primarily through antibody responses. Marzi *et al*. found that this vaccine generates a robust immune response in macaques to a West African strain of Ebola virus within days of immunization (see the Perspective by Klenk and Becker). Innate immune responses developed in as little as 3 days and increased the chances of survival, with complete antibody protection acquired 7 days after immunization. — CA

Science, this issue p. 739; see also p. 693

MICROTECHNOLOGY

CRISPR-Cas9 delivery by microfluidics

Delivering plasmids to difficult-to-transfect cells just got a bit easier. Han *et al*. report a microfluidic approach to deliver plasmids encoding single-guide RNA and Cas9 with high efficiency and high cell viability to mammalian cells. Rapid mechanical deformation of cell membranes generated temporary holes, permitting the transfer of plasmids without the use of viruses. The approach allowed for successful genome editing and gene loss-of-function in different cell lines across humans and mice. Editing mutations in this way could become an important tool for developing gene therapies. — PLY

Sci. Adv. 10.1126.sciadv.00454 (2015).

EVOLUTION

Helping the next generation diversify

Parasitism, including infections, can negatively affect fitness. Parents can help the next generation by increasing genetic diversity so that offspring can avoid or fight off these deleterious interactions more easily. For fruit flies, Singh *et al*. observed that in response to bacterial infection or predation by a parasitic wasp, the next generation showed increased recombination. However, this increase in genetic diversity was not due to increased recombination rates, but rather an unequal allocation of gametes that have undergone recombination. Infection therefore drives plasticity in the parental gametes, resulting in more diverse offspring. — LMZ

Science, this issue p. 747

GEOPHYSICS

Lower crustal deformation takes a turn

Collisions creating mountain belts frequently involve a tectonic plate plunging into the mantle. Huang *et al*. connect the deformation of rock from

IMAGES: (TOP TO BOTTOM) V. ALTOUNIAN/*SCIENCE*; NOBUYUKI MATSUDA/NTT BASIC RESEARCH LABORATORIES

Published by AAAS

the subducting plate to the surface topography in Taiwan (see the Perspective by Long). Subsurface deformation mapping required interpreting certain seismic wave velocities as they travel through the crust. The subsequent images of Taiwan's deep crust show two distinct layers of deformation. The bottom layer comprises the subducting slab, which is being pulled into the mantle. This mechanically couples with the upper layer of crust, compressing it into a mountain range. — BG

Science, this issue p. 720; see also p. 687

2D MATERIALS

Tuning the band gap of black phosphorus

Most materials used in electronics are semiconductors. The sizable energy gap in their electronic structure makes it easy to turn the conduction of electricity on and off. Graphene naturally lacks this band gap unless it undergoes certain modifications. Kim *et al.* studied the electronic structure of black phosphorus—a related two-dimensional material. By sprinkling potassium atoms on top of single layers of black phosphorus, the material changed from being a semiconductor to having a gapless linear dispersion similar to that of graphene. — JS

Science, this issue p. 723

LANGUAGE DEVELOPMENT

Marmosets learn to talk baby-talk

As human infants grow, their vocalizations change from cries, to babbles, to words. This pattern has been presumed to be absent from other primates. Indeed, the development of bird song is often regarded as a closer approximation of human language development. Takahashi *et al*., however, observed that marmoset cries and calls in the first 2 months after birth mature in much the same way as they do in humans (see the Perspective by Margoliash and Tchernichovski). Calls changed as the infants' vocal structures grew and were influenced by feedback from their parents. — SNV

Science, this issue p. 734; see also p. 688

DNA REPAIR

How to repair broken replication forks

Double-strand breaks in DNA are extremely dangerous to the integrity of our genomes. Most arise from problems encountered by replication forks during duplication of genomic DNA. Break-induced replication is known to use an error-prone DNA polymerase to repair such damage. Mayle *et al*. show that cells limit error-prone DNA synthesis by preventing the DNA polymerase from inadvertently switching to a related sequence with an incorrect template. The repair of the break is achieved by using a structure-specific nuclease to prevent formation of a long single-stranded region. — GR

Science, this issue p. 742

DRUG DELIVERY

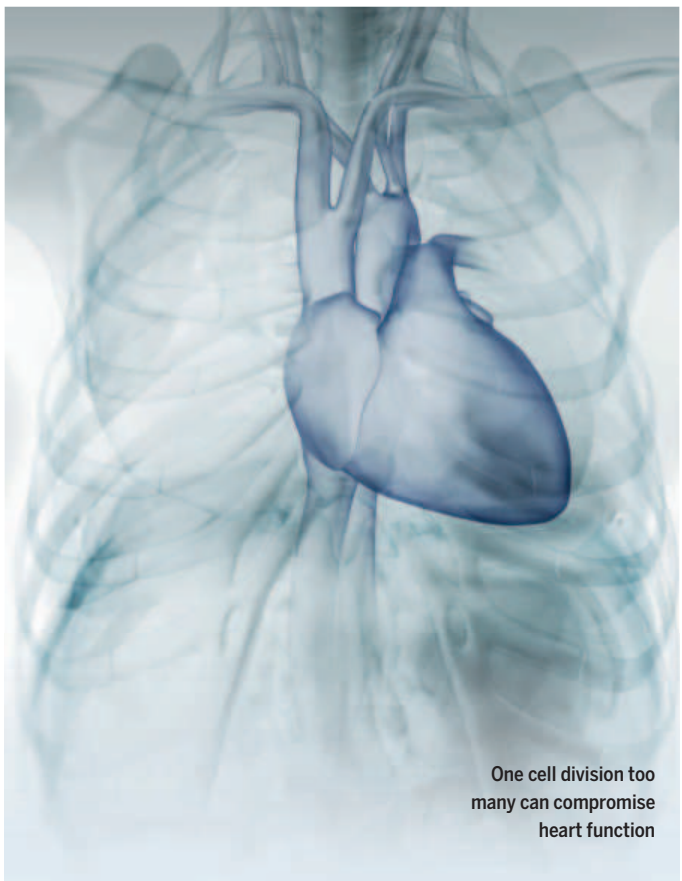
Hydrogels cozy up to inflamed tissues

Inflammation drives many chronic conditions. Directing potent drugs to the site of inflammation is highly desirable for improving treatment. Zhang *et al.* designed a hydrogel that self-assembles and delivers hydrophobic anti-inflammatory drugs directly to inflamed colon cells. Dexamethasone-loaded hydrogel enemas administered to a genetic mouse model of ulcerative colitis—a type of inflammatory bowel disease—relieved inflammation more effectively than free dexamethasone. In tissue samples from these patients, as well as in a chemically induced mouse model of colitis, hydrogel microfibers preferentially attached to inflamed tissue. — MLF

Sci. Transl. Med. **7**, 300ra128 (2015)

IN OTHER JOURNALS

Edited by **Kristen Mueller** *and* **Jesse Smith**



One cell division too many can compromise heart function

HEART DEVELOPMENT

The way to a broken heart

Mice lacking cardiac myosin binding protein C (MYBC) develop defective hearts that are twice the normal size. MYBC is a component of contractile thick filaments in the cardiac muscle. Jiang *et al*. found that the heart cells in mice lacking MYBC divided one extra time shortly after birth—when normal mouse heart cells would have stopped dividing. This caused the mice to have more myocytes with single nuclei—which compromise heart function. This unanticipated role of a structural protein in regulating how muscle cells divide may be important in different types of cardiomyopathy in human patients. — SMH

Proc. Natl. Acad. Sci. U.S.A. **112**, 9046 (2015).

NEUROSCIENCE

Out-of-register axons control output

A neuron integrates synaptic inputs and fires action potentials from its axon initial segment (AIS), a specialized membrane region on neuronal axons that also forms synapses with other axons. Stimulating neurons chronically can cause the AIS to move distally along the axon, but do its associated synapses move with it? To find out, Wefelmeyer *et al*. used optogenetics and imaging to show that in the rat hippocampus, synapses of chandelier interneurons on pyramidal neurons do not move with the

ILLUSTRATION: SCIENCE PICTURE CO/SCIENCE SOURCE

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Nick Wigginton**

GLIAL CELL SIGNALING

Cell type–specific glial networks

Glial cells respond to neurotransmitters when nerve cells communicate with each other. Glial cells themselves release gliotransmitters that regulate neural synaptic transmission. Martín *et al*. studied this reciprocal relationship in a brain region called the dorsal striatum, which has two types of experimentally identifiable neurons and two types of synapses (see the Perspective by Gittis and Brasier). Subpopulations of glial cells selectively responded to the activity of one specific type of neuron. In turn, these specifically activated glial cells signaled only to the same type of neurons but not the other, indicating that glial-nerve signaling is largely cell-type specific. — PRS

Science, this issue p. 730; see also p. 690

DEVICE TECHNOLOGY

Improving transistors with nanomaterials

High-performance silicon transistors and thin-film transistors used in display technologies are fundamentally limited to miniaturization. Incorporating nanomaterials—such as carbon nanotubes, graphene, and related two-dimensional materials like molybdenum disulfide—into these devices as gate materials may circumvent some of these limitations. Franklin reviews the opportunities and challenges for incorporating nanomaterials into transistors to improve performance. Because high-performance transistors are distinct from thin-film transistors, incorporating them into flexible or transparent platforms raises new challenges. — PDS

Science, this issue p. 704

HIV-1 VACCINES

Microbiota can mislead antibodies

Unlike the response to many viral infections, most people do not produce antibodies capable of clearing HIV-1. Non-neutralizing antibodies that target HIV-1's envelope glycoprotein (Env) typically dominate the response, which is generated by B cells that cross-react with Env and the intestinal microbiota. Williams *et al*. analyzed samples from individuals who had received a vaccine containing the Env protein, including the gp41 subunit. Most of the antibodies were non-neutralizing and targeted gp41. The antibodies also reacted to intestinal microbiota, suggesting that preexisting immunity to microbial communities skews vaccine-induced immune responses toward an unproductive target. — KLM

Science, this issue p. 705

PALEOCEANOGRAPHY

Slow circulation in the cold Arctic

The Arctic Ocean and Nordic Seas together supply dense, sinking water to the Atlantic Meridional Overturning Circulation (AMOC). The redistribution of heat by the AMOC, in turn, exerts a major influence on climate in the Northern Hemisphere. Thornalley *et al*. report that during the last glacial period, those regions were nearly stagnant and supplied almost none of the water that they presently contribute to the AMOC. This low rate of flow into the Atlantic was probably due to an absence of vigorous deep-water formation in the Arctic Mediterranean as a consequence of the extensive ice cover there at that time. — HJS

Science, this issue p. 706

APPLIED OPTICS

Chirality from achiral structures

The most common materials used in electronics produce only a simple optical response. Dhara *et al*. observed a complex circular photogalvanic effect in silicon nanowires, with the magnitude and direction of the induced photocurrent dependent on the polarization of the light. The specifics of the structure and geometry of the component materials are responsible for the effect. It should therefore be possible to engineer the same effect in other achiral materials and thus expand the box of enhanced functional materials for optical applications. — ISO

Science, this issue p. 726

MAGNETISM

Control of magnetism in heterostructures

The interface between two different materials in a heterostructure can exhibit properties unique to either of the two materials alone. A well-known example is a conducting gas that forms when $LaAlO_3$ is grown on $SrTiO_3$, but only if the $LaAlO_3$ layer is at least four unit cells thick. Wang *et al*. report a similarly abrupt magnetic transition in a heterostructure formed by another oxide ($LaMnO_3$) on the same $SrTiO_3$ substrate. Even though bulk $LaMnO_3$ is an antiferromagnet, when six or more unit-cell layers of it were deposited on $SrTiO_3$, it behaved like a ferromagnet. — JS

Science, this issue p. 716

CLIMATE CHANGE

Has there been a global warming hiatus?

Between 1999 and 2013, average temperatures at Earth's surface rose more slowly than in the preceding decades, despite fast rises in greenhouse gas emissions. This "hiatus" has led some to question the accuracy of climate model projections. In a Perspective, Trenberth explains how natural climate variability is likely to be larger on decadal time scales than commonly appreciated. Interannual and decadal processes in the Pacific Ocean strongly influence how fast temperatures rise. Although the underlying climate warming trend continues, natural climate variability can overwhelm it for short periods of time, as during the recent "hiatus." — JFU

Science, this issue p. 691

VASCULAR BIOLOGY

Maintaining vascular health with HDL

Flow through blood vessels subjects endothelial cells to abnormal shear forces at specific locations. This triggers inflammation, which contributes to atherosclerotic plaque formation. Galvani *et al*. found that the endothelial cell receptor $S1P_1$, which is activated by a lipid mediator abundant in blood, suppresses vascular inflammation and atherosclerosis in mice. This mediator, S1P, binds to different chaperone proteins and suppresses inflammation in cultured endothelial cells only when bound to the lipoprotein $ApoM^+HDL$. Thus, S1P may contribute to the protective effect of HDL—commonly called "good cholesterol"—in atherosclerosis. — WW

Sci. Signal. **8**, ra79 (2015).

Published by AAAS

REVIEW SUMMARY

DEVICE TECHNOLOGY

Nanomaterials in transistors: From high-performance to thin-film applications

Aaron D. Franklin

BACKGROUND: Transistors are one of the most enabling "hidden" technologies of all time and have facilitated the development of computers, the Internet, thin mobile displays, and much more. Silicon, which has been the material of choice for transistors in nearly every application for decades, is now reaching the fundamental limits to what it can offer for future transistor technologies. The newest display technologies are already turning to metal oxide materials, such as indium gallium zinc oxide (IGZO), for the improvements needed to drive organic light-emitting diodes. Ranging from applications such as display backplanes to high-performance microprocessors for servers, nanomaterials offer lasting advantages for the coming decades of transistor technologies. In this Review, the advantages of nanomaterials are discussed in the context of different transistor applications, along with the breakthroughs needed for nanomaterial transistors to enable the next generation of technological advancement.

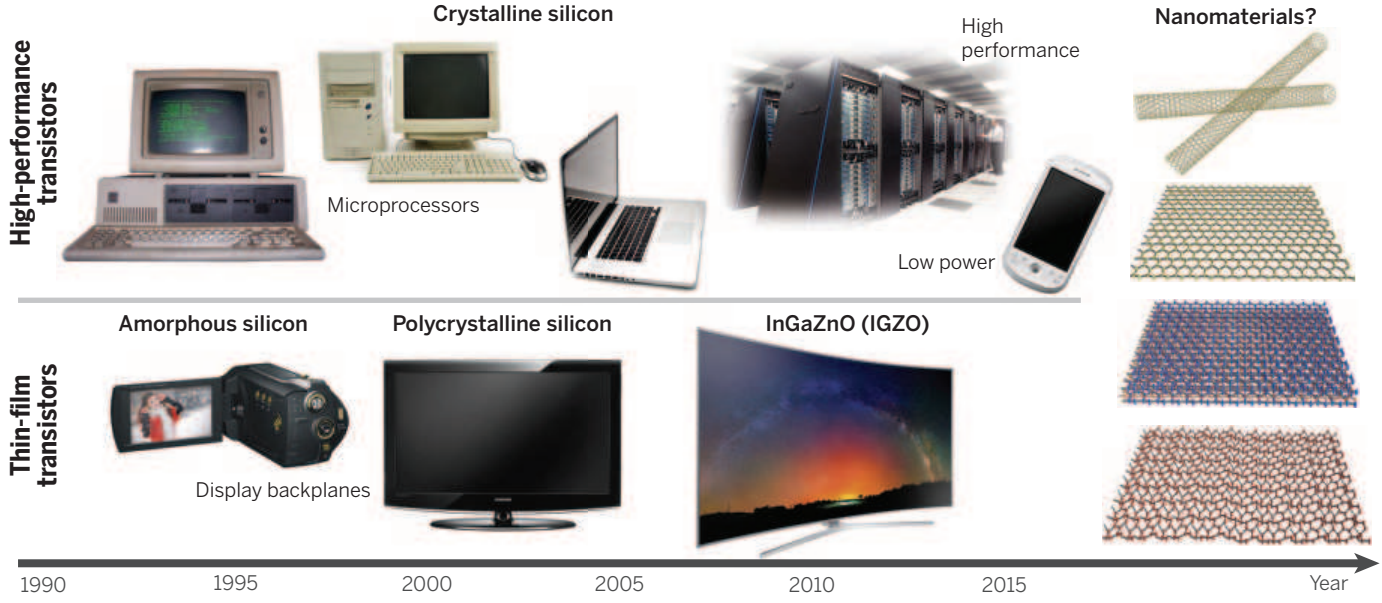
ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aab2750

ADVANCES: About 15 years ago, nanomaterials began receiving focused attention for transistors. Carbon nanotubes—molecules consisting of a cylindrical single layer of carbon atoms arranged in a hexagonal lattice—were the first to be given serious consideration, and their benefits quickly became widely acclaimed. Given their ability to transport electrical current with near-zero resistance, even at room temperature, the explosion of interest in nanotubes for electronics was understandable. Graphene, a related allotrope of carbon, benefited from the expansive interest carbon nanotubes had created for nanomaterial electronics. Although graphene transistors eventually proved less viable for digital applications, owing to the absence of an energy band gap, the excitement over graphene ushered in a complete revolution of interest in similar two-dimensional materials. Now, transition metal dichalcogenides and the so-called X-ene family of nanomaterials (e.g., silicence, phosphorene) dominate the attention of the nanoelectronics community. Hardly a day goes by without a paper being published on some advancement related to the use of nanomaterials in transistors. Hence, this Review focuses on how to keep such progress in the proper context with respect to the target transistor application, as well as the consideration of nanomaterials for completely new application spaces.

OUTLOOK: The benefits and practicality differ for each nanomaterial, and varied amounts of progress have been made in considering each of them for transistors. In just a few short years, thousands of papers have been published on improving synthesis or demonstrating simple functions of the newer nanomaterials. However, reflection on whether their newness translates to actual superiority over other options is warranted. Clearly, all of the nanomaterial possibilities offer certain advantages for future transistor technologies, but some do so with fewer caveats than others. Future research will benefit from keeping scientific advancement of nanomaterial transistors in line with end-goal deliverables. Overall, considering that only 15 years have elapsed since the study of nanomaterials for transistors began in earnest, the toolbox of available options and the developments toward overcoming challenges are promising. ■

Department of Electrical and Computer Engineering and Department of Chemistry, Duke University, Durham, NC 27708, USA. E-mail: aaron.franklin@duke.edu
Cite this article as A. D. Franklin, *Science* 349, aab2750 (2015). DOI: 10.1126/science.aab2750



Technologies enabled by high-performance and thin-film transistors over the past 25 years. (**Top**) Silicon transistors have driven the microprocessors used in computational devices ranging from low-power gadgets to large servers. (**Bottom**) Various forms of cheaper silicon enabled the display revolution, now being shared by IGZO. (**Right**) Nanomaterials may be the next transistor material for enabling a new generation of technologies.

CREDITS: OLD COMPUTER: WIKIMEDIA COMMONS/CREATIVE COMMONS; OTHER COMPUTER: JDRAY/FLICKR; LAPTOP: C. BERKELEY/FLICKR; SUPERCOMPUTER: WIKIMEDIA COMMONS/CREATIVE COMMONS; SMARTPHONE: PIXABAY/CREATIVE COMMONS; CAMERA: DAVJOY/FLICKR; FLAT SCREEN TV: AV HIRE LONDON/FLICKR; CURVED TV: SAMSUNGTOMORROW/FLICKR

REVIEW

DEVICE TECHNOLOGY

Nanomaterials in transistors: From high-performance to thin-film applications

Aaron D. Franklin

For more than 50 years, silicon transistors have been continuously shrunk to meet the projections of Moore's law but are now reaching fundamental limits on speed and power use. With these limits at hand, nanomaterials offer great promise for improving transistor performance and adding new applications through the coming decades. With different transistors needed in everything from high-performance servers to thin-film display backplanes, it is important to understand the targeted application needs when considering new material options. Here the distinction between high-performance and thin-film transistors is reviewed, along with the benefits and challenges to using nanomaterials in such transistors. In particular, progress on carbon nanotubes, as well as graphene and related materials (including transition metal dichalcogenides and X-enes), outlines the advances and further research needed to enable their use in transistors for high-performance computing, thin films, or completely new technologies such as flexible and transparent devices.

Fifty years ago, Gordon Moore published his seminal paper that included a projection that the semiconductor industry would double the number of components on a chip (integrated circuit) roughly every couple of years (*1*). Although slightly modified over time, this projection from one of Intel's founders has served as the marching orders for what is now a >$330 billion global industry (*2*) and has been given the moniker "Moore's law." Notable innovations have allowed the silicon metal oxide semiconductor field-effect transistor (MOSFET) to be made smaller, the latest being from strained-silicon channel materials to three-dimensional (3D) fin gate structures, as depicted in Fig. 1C. Despite these advances, fundamental physical limits for the minimum size of silicon MOSFETs are now being approached, and the question of "what's next" for transistors has become inescapable.

The microprocessor, now driven by billions of MOSFETs on a single chip, is the most prominent transistor application, as it is the computational "brain" to every electronic system. Yet there are other needs for the transistor that do not necessarily require the high performance (and high cost) of MOSFETs—for example, chemical and biological sensors, optical detectors, and the pixel-driving circuits for displays. Aside from the microprocessor, the most prominent use of transistors is in the backplane electronics of flat-panel displays. Departure from the cathode ray tube was made possible, in part, by the use of low-cost amorphous silicon (a-Si) thin-film transistors (TFTs) in liquid-crystal displays.

Department of Electrical and Computer Engineering and Department of Chemistry, Duke University, Durham, NC 27708, USA. **E-mail: aaron.franklin@duke.edu**

That a-Si TFTs were made from silicon is one of very few attributes TFTs have in common with silicon MOSFETs. MOSFETs are generally high-performance and high-cost transistors used for computational devices (e.g., smartphones, computers, servers), whereas TFTs are less cost intensive and are used in applications with much lower performance requirements. Consider the different perspectives on Moore's law shown in Fig. 1 for these two types of transistors. The traditional Moore's law in Fig. 1A is obeyed by high-performance transistors, wherein the density of devices doubles approximately every 2 years (*3*). For TFTs, the integration density has changed relatively little, with the focus being more on improving stability and driving down cost. The Moore's law perspective in Fig. 1B focuses on the component substrate size (*4*); here, the trend for TFTs is based on the display market, where the ability to fabricate large areas of transistors is crucial (*5*). Meanwhile, the die size of the MOSFET—the area of a single chip on a much larger production wafer—has been capped at ~300 mm^2 to minimize production costs by yielding more chips per wafer. Overall, these two perspective plots highlight the importance of recognizing the differences in applications and deliverables between high-performance transistors and TFTs.

Nanomaterials are of great interest for use in transistors of all types, as they offer many electrical and mechanical advantages. Unlike silicon-based transistors, the general structure of a nanomaterial transistor changes very little between the high-performance and thin-film varieties, except in size (see Fig. 1C). This similarity has led to much confusion in the research community as to whether certain reported nanomaterial transistor advancements are useful for high-performance or thin-film applications. Such confusion occurs only for nanomaterials, as the silicon community developed different materials expressly for TFTs. Consider the substantial difference in target performance metrics—often orders of magnitude—between high-performance transistors and TFTs (from any material), as shown in Table 1. In this Review, in addition to considering the great progress and promise in the field of nanomaterial transistors, clarification of this important distinction between the intended application spaces will be provided. Guided by an understanding of the relevant dimensional and performance targets (see Table 1), it is hoped that future research on nanomaterial transistors can more accurately address the scientific challenges limiting their ultimate realization.

Nanomaterials

The term "nanomaterial" can refer to any material with dimensions at the nanoscale (<100 nm), but for transistor applications, the materials that naturally exhibit quantum confinement and occur without the need for thinning or patterning are of greatest interest. Hence, the focus of this Review will be on such naturally quantum-confined nanomaterials, including single-walled carbon nanotubes (CNTs), graphene, transition metal dichalcogenides (TMDs), and X-enes (the newest and least naturally occurring variant of 2D crystals, typically consisting of a buckled hexagonal structure to mimic graphene—hence the name "X-ene"; examples include phosphorene for phosphorous and silicene for silicon). There certainly are benefits in transistors from other types of nanomaterials, such as nanowires (*6*, *7*), which are nanoscale versions of bulk materials. However, it is typically best to avoid quantum confinement in such materials rather than embrace it as a key aspect of their electronic structure because it would introduce considerable variation in the resulting device performance and degrade carrier transport properties. Most of the nanomaterials discussed herein are often classified as van der Waals materials, as they do not covalently or ionically bond to other materials but exist as 1D or 2D constituents that are either isolated or assembled in some fashion via van der Waals weak attraction forces.

Extensive articles reviewing the distinctive properties of each of these nanomaterials have been published (*8*–*15*); as the focus of this Review is on the use of the nanomaterials in transistors, only a brief summary of their key attributes is given. Of note is that single-walled CNTs and graphene share the sp^2-bonded, hexagonal carbon lattice and thus exhibit similar carrier transport properties, including high Fermi velocity (v_F), which can lead to higher switching speeds. Low effective mass and correspondingly high carrier velocity make CNTs and graphene the most ideal electron transport systems available. Consequently, CNTs are favorable options for very small transistors that can operate at low voltages, thus conserving power, because silicon-based transistors suffer degradation in

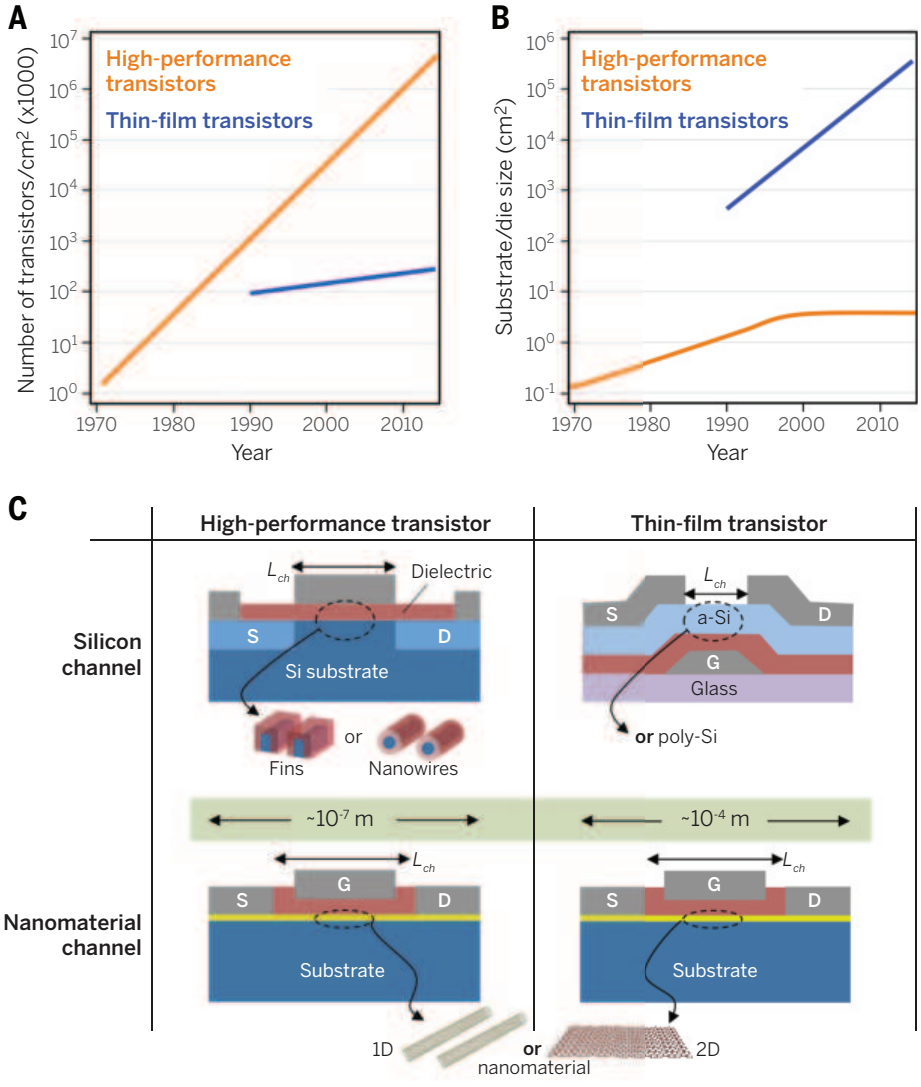


Fig. 1. High-performance transistors versus TFTs: Moore's law and device comparison. (**A**) Approximate plot of Moore's law, showing increasing density of integrated transistors (reduction of transistor size). High-performance transistors exhibit the traditional trend, with a density that doubles approximately every 2 years, whereas there has been little change in the approximate density of TFTs. (**B**) A different view of Moore's law, showing the approximate substrate or die size for the two transistor technologies over time (TFT substrate size is based on display applications and has doubled every ~2.5 years). For (A) and (B), lines are approximate fits to actual data, as in (*3*, *4*). (**C**) Cross-sectional schematics comparing the general structure of high-performance transistors and TFTs with silicon versus nanomaterial channels. Note that the latest high-performance silicon device could consist of fins (3D) or nanowire channels, as shown. Also indicated is the channel length L_{ch} for each device. In the nanomaterial channel transistors, where the channel is made up of 1D or 2D nanomaterials, there is little difference in structure between high-performance transistors and TFTs, yet size differs by orders of magnitude—a key source of confusion in the field. S, source; G, gate; D, drain.

performance (e.g., leakage currents) when operated below ~1 V. The nanotubes, which can be conceptualized as rolled up sheets of graphene, are circumferentially quantum-confined, which makes them true 1D electronic systems with reasonable energy band gaps that allow them to turn off with little leakage current. The lack of an energy band gap in graphene is the nanomaterial's greatest pitfall for transistors, rendering the devices unable to block current flow and be switched to the off state (digital "0"). Great efforts have been made to induce a band gap in graphene (*16*, *17*), but to date, they lack practicality. Even thinning graphene to create quasi-1D nanoribbons (*18*, *19*) to create a band gap still has disadvantages, such as edge states, compared with the naturally 1D nanotubes. The most likely use of graphene in digital transistors arises from its ability to enable distinct 2D heterostructures, which will be discussed below. A summary of some of the most relevant intrinsic properties of these nanomaterials for electronics applications is given in Fig. 2. Note that the focus in Fig. 2 is not given to mobility or extrinsic device metrics, such as contact resistance, as these are only relevant in certain devices or are highly variable, depending on specific interfacial materials. Further discussion of such other metrics is given below.

The interest in studying graphene for transistors spurred reconsideration of a very well-known class of van der Waals nanomaterials that had not been studied for electronics in many years: TMDs. Many combinations of transition metals (e.g., Mo, W, Hf) and chalcogens (e.g., S, Se, Te) can yield the three-atom-thick arrangement of a monolayer TMD (*9*). For transistors, the greatest interest has been in the Mo and W families, especially MoS_2. Electron transport in TMDs is slower by a factor of 20 compared with carbon nanomaterials, but TMDs offer sizable energy band gaps for switching and maintain the attractive 2D confinement of graphene. If more than one monolayer is stacked to form a many-layer TMD, the band gap changes markedly, typically saturating to approximately the bulk value at ~15 layers with a gap that is barely two-thirds that of the monolayer.

Most recent to join the options of nanomaterials for transistors is the so-called X-ene family (*20*–*26*). Exploration began a few years ago, motivated by the electronic structure of graphene, including a linear dispersion relation of Dirac cones, for other group IV and V nonmetals arranged in a similar fashion to the graphene hexagonal lattice. Thus far, demonstrations of silicene (silicon) (*21*), germanene (germanium) (*24*), phosphorene (black phosphorus) (*23*, *25*, *26*), stanene (tin) (*20*), and arsenene (arsenic) (*22*) have been made. Their lattice structures are not as perfectly planar hexagonal as that of graphene but tend to be some variation in a buckled hexagonal form (*20*). Phosphorene exhibits van der Waals stacking, but the other X-enes are synthesized as adlayer structures on certain substrates. Limited experimental work has been performed on X-enes, but theoretical projections indicate potentially attractive electronic properties, including the presence of a reasonable band gap and transport behavior about half as favorable as that of graphene (Fig. 2). One of the biggest challenges for X-enes compared with other nanomaterials is that X-enes tend to be highly reactive in air, making even simple device structures extremely difficult to realize (*21*).

To consider the progress that has been made in demonstrated nanomaterial transistors, a summary of a few key device metrics is given in Fig. 3A. For high-performance transistors, there is a need to scale the channel length L_{ch} and contact length L_c while operating at low voltage V_{DD}. Note that the approximate progress shown for each metric in Fig. 3A is from diverse reports; in other words, there has yet to be a solution that shows the scaling of all relevant dimensions along with low-voltage operation. As will be discussed below, the focus for TFTs is on mobility μ and on-state current I_{on}, where nanomaterials have already shown substantial improvements over silicon-based options. Based on the Fig. 3A comparisons, CNTs show the most consistent promise thus far,

which is partly an artifact of their being studied the longest but also undoubtedly related to their superior carrier transport properties compared with the TMDs or X-enes. These metrics are certainly not comprehensive, but when considering any of these nanomaterials for the indicated applications, such target deliverables must be kept in mind.

Benefits of nanomaterials

Despite the substantial differences in intrinsic properties among the nanomaterial options, they have common advantages for transistors, as summarized in Fig. 3B. Foremost, nanomaterials are all atomically thin. For high-performance transistors, there is a need to scale the channel length to sub–10-nm dimensions because this will be required for technologies targeted for the early 2020s, and silicon is unable to operate at such length scales. Such scaling requires the transistor gate to maintain electrostatic control over the channel—a difficult feat when the channel thickness (d_{body}) is greater than L_{ch}. The atomic thinness of nanomaterials, especially in their monolayer form for 2D nanomaterials, offers ideal electrostatic control. Often the ultimate scalability of a transistor with an undoped (intrinsic) channel—such as all nanomaterials shown in Fig. 2—is determined by the screening or natural length λ being proportional to d_{body}. The rule-of-thumb is that a $L_{ch} \geq 3\lambda$ will ensure that the gate maintains electrostatic control and that deleterious short-channel effects are avoided (*27*).

Another benefit shared by most nanomaterials is their substrate independence. CNTs can be synthesized in a reactor chamber, dispersed in solution to isolate a certain diameter or energy band gap (*28*–*38*), and then deposited onto any substrate of choice, all while keeping their crystal structure and transport properties intact. The same is true for graphene and TMDs, with the exception of X-enes, as they are generally substrate-bound (except for phosphorene). Although synthesis conditions are extreme in most cases, the ability to transfer nanomaterials to virtually any substrate and fabricate transistors from them is one of their greatest strengths.

Heterostructures of the nanomaterials in Fig. 2 are becoming an attractive possibility for transistors, including the use of all 2D nanomaterials for fabricating transistors (e.g., graphene contacts, TMD channel, and hexagonal boron nitride gate dielectric) (*39*–*41*), as well as the stacking of diverse TMDs in a fashion similar to epitaxial III-V materials (*42*–*49*). Work continues on understanding what governs transport at these nanomaterial heterostructure junctions. Several devices that make explicit use of such heterostructures in their operation have been demonstrated, and it will be exciting to see the types of new transistors this approach will enable.

Other strengths that nanomaterials offer for transistors typically depend on device structure or type. For instance, the presence of a small quantum capacitance (dependent on the density of states), especially in CNTs, has been shown to enable more extensive modulation of the surface potential in the on state of a transistor (*50*). Operation in this quantum capacitance limit is of great value for tunneling transistors, for which the thinning of tunnel barriers to achieve higher current is crucial. Other nanomaterial properties that are useful for specific devices include spin transport, Coulomb blockade, charge density

Table 1. Comparison of key target metrics for high-performance transistors and TFTs. Channel length is defined in Fig. 1 and is critical for the packing density of transistors. I_{on} indicates the on-state performance, including transistor speed. V_{DD} reflects the off-state performance, with lower V_{DD} needed for densely integrated high-performance applications. Mobility is discussed below and has become relevant only for TFTs. Process complexity is given in terms of the number of mask layers or levels used to fabricate the transistors. Although this number will vary for high-performance transistors, depending on the application (digital logic versus memory, server chip versus low-power system-on-chip, etc.), the approximate average is given and clearly shows the advantage for TFTs with <20% the process complexity. Cost for fabrication line or facility (fab) is based on the most recent facilities reported—and the numbers for TFTs are for fabs that produce full OLED displays (not just the TFT backplane)—yet it is still considerably cheaper than a high-performance transistor fab. The dash indicates no ability to numerically compare.

Metric	High-performance transistors	Thin-film transistors	Difference
Channel length (L_{ch})	<20 nm	5 to 10 μm	250×
Drive current (I_{on})	>1 mA/μm	>10 μA/μm	100×
Operating voltage (V_{DD})	<1 V	3 to 5 V	3× to 5×
Mobility (μ)	see Fig. 4	0.1 to 100 cm^2/Vs	–
Process complexity	>40 mask layers	<7 mask layers	5.7×
Cost for fab	$4 billion to $10 billion	$1 billion to $3 billion	3× to 4×

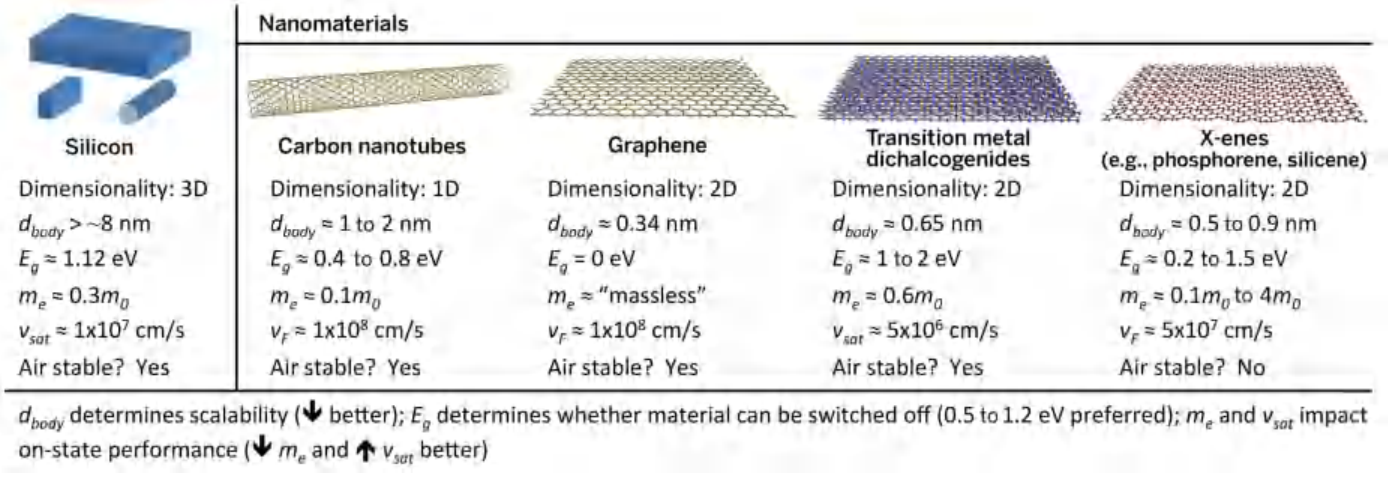


Fig. 2. Nanomaterial options for transistors. A comparison of key intrinsic attributes of the four most prominent nanomaterials (or families of nanomaterials) is shown contrasted with the related attributes of silicon. Carbon nanotubes and graphene share the sp^2 bond structure and thus many electrical transport properties, with the exception that the nanotubes are circumferentially quantum confined to form 1D cylinders with band gaps. The TMDs that are most amenable for transistors are the Mo and W families (e.g., MoS_2, WSe_2). X-enes, typically consisting of a buckled hexagonal structure and lacking stability in air, are the newest and least naturally occurring variant of 2D crystals. The channel or body thickness (d_{body}) limit for silicon is based on the onset of quantization effects that must be avoided for a bulk material. E_g, band gap; m_0, electron rest mass; v_{sat}, saturation velocity; v_F, Fermi velocity.

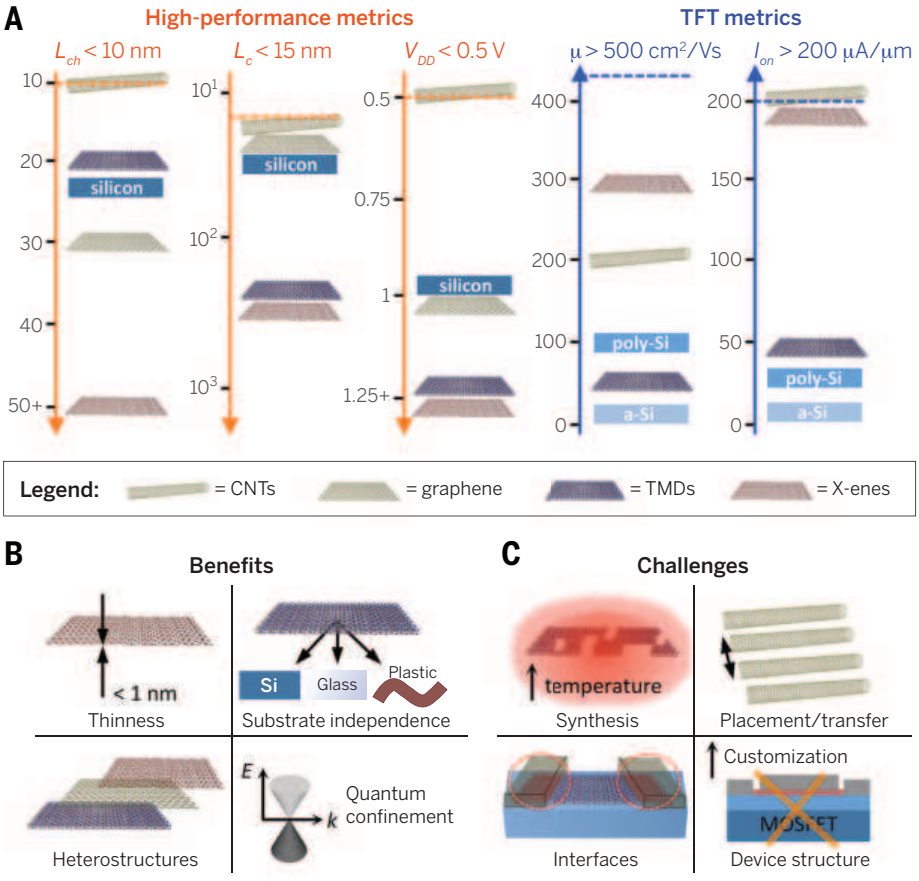


Fig. 3. Performance comparison, benefits, and challenges for nanomaterials in transistors. (**A**) Approximate comparison of demonstrated performance or device metrics, where L_{ch} is channel length and L_c is contact length (length over which the metal contact covers the nanomaterial). Note that the approximate values for the nanomaterials are from reports involving transistors with relevant dimensions and structure, and the metrics would need to be simultaneously met (along with others not included) for the targeted transistor technology. Also shown are the approximate values for the most advanced transistors from silicon-based materials. (**B**) Overview of transistor benefits offered by all nanomaterials, including atomic thinness, substrate independence, potential for heterostructures, and distinctive electronic properties. *E*, energy; *k*, wave number. (**C**) Overview of challenges shared in some fashion by all nanomaterials for transistors, including high-quality synthesis, controlled placement, improved interfaces (contact and dielectric), and devices that more fully utilize nanomaterial advantages.

waves, and plasmons, but the focus of this Review will be on more conventional digital transistors.

Challenges for nanomaterials

Just as the nanomaterials in Fig. 2 share many advantages for transistors, they also have challenges in common, as summarized in Fig. 3C. Without question, the biggest challenges relate to the controlled synthesis and placement of the nanomaterials. Although there has been widely varied progress on addressing the synthesis and placement challenges among the different nanomaterials, they all require substantial improvement to be viable for a transistor technology. Synthesis of nanomaterials takes place at high temperatures (typically >800°C). In most cases, the nanomaterials are synthesized on a sacrificial substrate and then transferred in some fashion to the substrate for fabricating transistors. Whether or not they are kept on their synthesis substrate, the distribution in d_{body} (diameters for CNTs and number of layers for 2D nanomaterials) must also be controlled. Tremendous progress has been made in separating CNTs of certain electronic type (band gap in CNTs is dependent on d_{body}) (*28*–*38*), which indicates promise for reaching technology targets so long as scientific investment continues (current highest verified purity is 99.99% with a target of 99.9999%) (*51*). Progress has also been made in synthesizing TMDs (*52*–*60*), although there has yet to emerge a process capable of growing a TMD with complete coverage, high quality (low defect and grain boundary density), and a uniform number of layers. Being the newest explored nanomaterial and not exhibiting van der Waals stacking behavior, X-enes have the farthest to go for improved synthesis.

If the nanomaterial is synthesized on one substrate and then transferred to another for transistor fabrication, then precision placement is important. For CNTs, this problem is especially pronounced, as they require accurate positioning into arrays. The target pitch for CNTs in high-performance transistors is 5 to 8 nm (125 to 200 CNTs per micrometer) (*13*), and promising advancements continue to be made to this end (*61*–*68*), including the use of selective deposition to predefined regions of a substrate (50-nm pitch achieved). It is less favorable to have tightly packed (~0-nm pitch) arrays of CNTs for high-performance transistors (*62*), as they will result in deleterious charge screening effects and challenges for establishing good electrical contacts. Requirements for TFTs are much more relaxed, where even tangled films of CNTs with no alignment can be used (*69*–*71*). The 2D nanomaterials do offer planar coverage, but it is difficult to transfer the films without inducing defects that are detrimental to transistor performance. An additional complication for X-enes is their pronounced dependence of carrier transport on crystallographic direction (*23*); orientation of the X-ene film will change the effective mass of the carriers by up to one order of magnitude.

Interfaces to nanomaterials are another challenge. Without having available surface states in the way that bulk materials do, the formation of covalent bonds to nanomaterials can be difficult. For instance, nucleating the growth of high-quality dielectrics (insulating barriers) with atomic layer deposition (ALD) is problematic for nanomaterials, as they do not offer typical end groups for reacting with the ALD precursors. Creative solutions have been presented for potentially addressing the creation of high-quality dielectric interfaces (*72*–*75*), but there has been much less progress on improving the contact metal interfaces. Regardless of whether a nanomaterial transistor is for high-performance or thin-film applications, the device will depend heavily on the quality of transport at the source and drain metal contact interfaces. Studies have been performed on the impact of different metals on the contact resistance for CNTs (*76*–*78*), TMDs (*79*–*81*), and even X-enes (*82*), most of which consider the metal-nanomaterial interface as a traditional Schottky barrier structure. However, with no covalent bonding between the metal and nanomaterial, the reality of what determines transport at the interface remains elusive, and the Schottky barrier picture has been shown to break down under certain conditions (*78*).

In all cases, the metal-nanomaterial contact interface requires further scientific study and engineering improvement to yield the necessary performance and consistency for a transistor technology. Many have interpreted this to simply mean the achievement of low contact resistance (R_c), but depending on the transistor application, realizing a low R_c may not address all of the relevant contact issues, including scalability and reproducibility. For example, for high-performance transistors with contact lengths <15 nm (see Fig. 3A), the target R_c must be <150 ohm·μm (measured contact resistance multiplied by device width) per contact (*78*)—a

substantial challenge even at very long (>200-nm) contact lengths.

A final point regarding challenges for nanomaterials in transistors is the need for further discovery and innovation of the transistor structures. Thus far, nanomaterials are integrated into transistors that essentially mimic the silicon MOSFET, especially for high-performance applications (Fig. 1). Considering their substrate independence, there are undoubtedly much more favorable transistor structures that have yet to be discovered for nanomaterials. Such structures would take more specific advantage of the electrical and mechanical properties of nanomaterials for boosting performance, lending new functionality, or both for certain applications. Additionally, regardless of the transistor structure into which nanomaterials are integrated, a need remains for studying the reliability of resultant devices and circuits to ensure their utility under certain application requirements, such as high clock speeds.

High-performance transistors

High-performance transistors are those that enable the primary computing electronics that we rely on daily. Servers that fill warehouses (server farms) are the lifeblood of the Internet and rely on the highest-performing scaled transistors. Semiconductor companies such as Intel and Taiwan Semiconductor Manufacturing Company will develop the transistor technology—for example, "14 nm" (*83*)—for the highest-performance applications and then use mostly the same transistor for other less-demanding applications. An example is the low-power chips that are used for mobile electronics (e.g., smartphones, tablets, laptops), where the chip will still make use of transistors with the latest technology (14 nm) but will be integrated less densely and run at lower voltages. In this case, the term "high-performance transistor" refers to the transistor technology itself that is applied to all such peripheral applications—from servers to smartphones.

A closer look at Moore's law for high-performance transistors, given in Fig. 4, helps to highlight why nanomaterials are so advantageous. In the early 2000s, the performance (as measured by chip clock frequency) was necessarily capped to address the runaway power density problem (*84*): Smaller transistors could leak power even in the off state, leading to power-consumption and heat-generation issues. Anything beyond ~100 W/cm^2 is detrimental to the chip and surrounding components; heating a semiconductor causes performance fluctuations, and the ability for cooling technologies to dissipate the heat is limited to ~100 W/cm^2. Active power (P_A) depends on clock frequency (f), operating voltage (V_{DD}), and the number of transistors per square centimeter (N) as $P_A \propto N^* f^* (V_{DD})^2$. With Moore's law demanding that N continue to increase, the most logical solution would be to reduce V_{DD} because of the squared dependence. However, the cost in terms of performance was too great, and V_{DD} has remained pegged for more than a decade at ~1 V for high-performance transistors. Reducing f would also seriously compromise performance, leading to the need for multicore architectures. The advantage of multicore chips in terms of the power problem is that the architecture ensures that, at any given time, a much smaller fraction of the transistors will be turned to the on state than for a single core chip—a phenomenon known as dark silicon (*85*). However, the multiplication of the number of cores will hit a level of maximized benefit at ~16 cores, leaving reduction of V_{DD} the only hope for further progress.

Nanomaterials offer several ways of reducing the operating voltage in high-performance transistors. First, their atomic thinness enables highly effective gate control over the channel potential through the reduced screening length discussed above. Improved gate control means that even at aggressively scaled channel lengths (needed for driving an increase in N), nanomaterial transistors can switch to the on state at the lowest possible voltages (i.e., small subthreshold swing, which is an indicator of how many volts are required to switch the current in a transistor by one order of magnitude). The subthreshold swing is referred to as the performance metric in the off state, indicating the ability of a transistor to switch at low voltages. Beyond improved gate control, nanomaterials also enable advanced transistor structures that may offer solutions for scaling V_{DD}. One such option is the tunneling transistor, where the small effective mass (m_e) in CNTs (as well as some 2D options) would offer dramatic improvement (increase) in the tunneling current to boost the on state, which limits the realization of these transistors (*86*, *87*).

Exceptional electron transport in most nanomaterials is another attractive feature for high-performance transistors. In the case of CNTs, nearly ballistic (zero resistance) transport has been observed at room temperature at lengths up to 40 nm (*88*, *89*). Such favorable transport has experimentally been shown to enable sub–10-nm CNT transistors that outperform any similarly scaled silicon-based transistor at low voltages ($V_{DD} \leq 0.5$ V) (*90*). Transport in TMDs is actually worse than in most bulk semiconductors, including silicon; yet at sufficiently scaled channel lengths, TMDs could still offer advantages, as performance in the transistor will become dominated by contacts rather than transport through the channel. For the required dimensions of a high-performance transistor, the performance of every device will be determined more by the contacts than by the channel, regardless of whether it is a nanomaterial or silicon or any other semiconductor. For this reason, the use of the prevalent transport metric, mobility, must be reevaluated.

Mobility

One of the most widely used metrics for describing the quality of carrier transport through a semiconductor is mobility (μ). Initially defined as the constant of proportionality between carrier velocity and the electric field across a material, μ has often been considered an intrinsic indicator of semiconductor quality and even quoted in textbook tables. Though it is understood that μ depends on other factors such as doping density and temperature, it has nevertheless been taken as the most important parameter for semiconductor consideration. In fact, consideration of channel materials other than silicon (which suffers from a relatively low μ compared with other bulk semiconductors) for high-performance transistors has largely focused on semiconductors with higher mobility, eventually leading to the classification "high-mobility semiconductors" (usually germanium or III-V compound semiconductors, such as GaAs, which despite higher mobility have other issues in achieving high levels of integration).

For the first 50 years of Moore's law, using mobility as a key parameter to indicate the quality of a semiconductor for transistors made sense. Channel lengths were long enough to yield an average transport length between scattering events (mean free path) or time between scattering events (mean free time τ) and, thus, a certain mobility: $\mu = q\tau/m_e$ (here, q is elementary charge of an electron). What happens if the channel length falls below the average length between scattering events, or, in the most extreme case, what happens if the channel is ballistic? As shown in Fig. 5, the use of mobility in such a short-channel transistor is misleading, as it no longer accurately indicates the quality of carrier transport through the semiconducting channel.

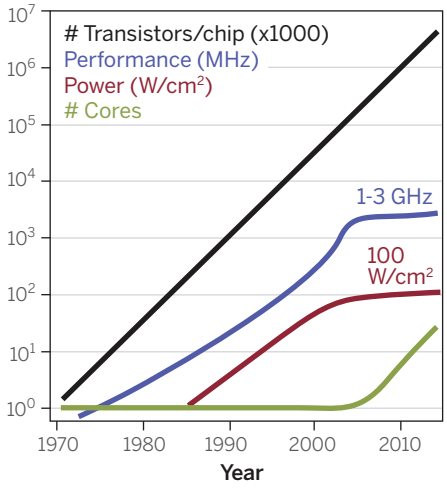


Fig. 4. Detailed look at Moore's law for high-performance transistors. Approximate plot of various metrics in high-performance microprocessors over the past >40 years [lines are approximate fits to actual data, see (*3*)]. The oft-quoted version of Moore's law is the number of transistors per chip (black line), but key trade-offs in chip performance (clock frequency, blue curve) have been made since the early 2000s because of the power density (red curve) reaching a physical limit of ~100 W/cm^2. Continued Moore's law scaling of high-performance transistors has been sustainable only by multicore architectures (green curve)—and the associated "dark silicon"—for the past decade.

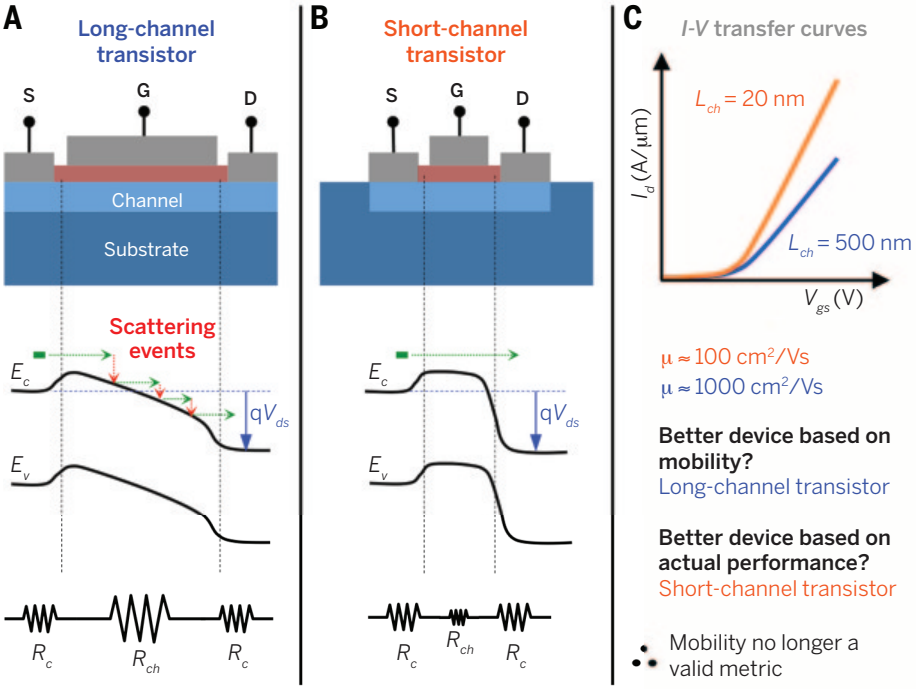


Fig. 5. Is mobility meaningful for all transistors? (**A**) Generic schematic of a long-channel transistor with a corresponding energy band diagram illustrating how the scattering events (red) of electrons (green) moving across the channel lead to a potential drop (sloped bands) of the applied drain-source field (qV_{ds}) and, hence, a substantial resistance in the channel (R_{ch}) compared with resistance at the contacts (R_c). E_c, conduction band edge; E_v, valence band edge. (**B**) Same as (A), but for a short-channel transistor where there are very few scattering events in the channel (quasi-ballistic) and hence the potential is dropped at, and the device is limited by, the contacts. (**C**) Conceptual transfer curves for the two devices, showing how mobility (μ) mistakenly suggests better performance in the long channel. This diagram illustrates why mobility is no longer a meaningful metric for short-channel (high-performance) transistors but is still valid in long-channel (thin-film) transistors. *I*, current; *V*, voltage; I_d, drain current.

With performance in short-channel transistors being determined almost exclusively by the injection of carriers at the source-drain contacts, extraction of mobility from device data defies the actual meaning of the parameter. The example in Fig. 5 illustrates how the band structure and relevant resistances change between long- and short-channel transistors. Extraction of mobility from two such transistors could yield μ that is an order of magnitude smaller for the short-channel devices, even though the performance is much better. This is a fallout of the widely used field-effect mobility expression, wherein $\mu = g_m L_{ch}/(WCV_{ds})$ (here, W is the device width, C is the capacitance, and V_{ds} is the drain-source voltage), where the transconductance (g_m) goes up but not nearly as much as L_{ch} goes down when scaling from long to short channels (all other terms remain the same). The contacts, including R_c, are identical in the two devices, as is the applied voltage, but the extraction of mobility from the device data relies on the assumption that transport is limited by the scattering in the channel, as in the long-channel case.

In reality, mobility is no longer meaningful for high-performance transistors (all of which are of the short-channel variety), regardless of whether they have silicon, III-V materials, or nanomaterial channels. The saturation velocity of carriers will matter, as will the effective mass, for determining how well carriers move through the channel. Consider the difference in common expressions for current in a long-channel (traditional) MOSFET (*91*)

$$I \approx \mu \frac{W}{L_{ch}} C_{ox}(V_{gs} - V_t)V_{ds}$$

versus a 1D short-channel nanomaterial transistor (*92*)

$$I \approx q\int f(E,T) \cdot v(E) \cdot T_p(E) \cdot D(E) \cdot dE$$

In addition to voltage dependencies [gate-source voltage (V_{gs}), threshold voltage (V_t), and V_{ds}], current in the long-channel transistor depends on mobility, gate capacitance (C_{ox}), and spatial parameters (W and L_{ch}). In contrast, the short-channel nanomaterial transistor shares none of these dependencies. Rather, the current for the 1D nanomaterial transistor relies on the Fermi function [f (indicates electron distribution)]; carrier velocity (v); transmission probability (T_p), including the probability that carriers will make it through the metal-nanomaterial contact without scattering; and density of energy states (D), with respect to energy (E) and temperature (T). The only material consideration in the long-channel expression is μ, whereas the short-channel transistor accounts for D, v, and f, as these are the material-related parameters that become relevant when scattering in the channel is minimal. Hence, at this point of development, relying on mobility as a metric when discussing options for high-performance transistors causes confusion and distraction from the parameters that matter most. This is especially true for nanomaterials, for which reports of mediocre mobilities extracted from short-channel transistors undercut the actual potential of the material. Mobility remains a useful parameter for comparing TFTs, as they are long-channel devices.

Thin-film transistors

There are many applications for which a digital transistor switch is needed without the extreme performance requirements of computational devices. This became especially clear in the 1980s, when a revolution was beginning to stir in display technology with the need for a more compact, mobile display for laptops. Liquid crystal displays were favored but needed transistors for the backplane to drive the pixels. In 1981, researchers showed that a-Si provided the answer—it could be deposited onto glass substrates, was mostly stable, and had sufficient mobility (0.1 to 1 $cm^2\ V^{-1}\ s^{-1}$) to yield the necessary level of drive current (I_{on}). The size of the a-Si TFTs did not have to be small, as the typical pixel is on the order of 100 μm by 100 μm, and the TFTs needed to be large to provide the required I_{on}. Perhaps most importantly, the a-Si TFTs could be fabricated on large substrates at relatively low cost.

With a-Si TFTs, the display revolution began. Shortcomings, including bias-stress instability (change in the voltage needed to turn the TFT on after prolonged operation) and low mobility, led to development of other TFT options, including polysilicon (μ ~ 100 $cm^2\ V^{-1}\ s^{-1}$) (*4*). Although too costly for replacing a-Si in the backplane, polysilicon has become important for display driver circuits in mobile applications, as it is still considerably cheaper than using high-performance MOSFETs. Much effort was devoted to using polysilicon to drive organic light-emitting diode (OLED) displays, which require much higher drive currents than a-Si TFTs could ever deliver, but the nonuniformity in threshold voltage in the polysilicon TFTs gave concern for causing nonuniformity in display brightness (*5*). The latest advancement has been the use of metal oxide materials—in particular, InGaZnO (IGZO), which offers manufacturing costs on par with a-Si, mobility near that of polysilicon, and better stability. IGZO TFTs now drive the latest OLED displays on the market, yet they have their own limits based on processing temperature and mobility pegged at ~100 $cm^2\ V^{-1}\ s^{-1}$ for reasonable synthesis temperatures.

Nanomaterials hold much promise for TFTs. Even though being formed into an inhomogeneous thin film seems to belie the advantages of the nanomaterials, they still are shown to deliver superb performance and processing benefits. For example, CNTs that are simply drop cast from

solution onto a substrate to form a thin film (the cheapest fabrication approach possible, as it is amenable for printing processes) consistently deliver mobilities of 10 to 100 $cm^2 V^{-1} s^{-1}$ (*69*–*71*, *93*–*95*). Further efforts to induce alignment in the CNT thin films could boost mobility much higher than 100 $cm^2 V^{-1} s^{-1}$ (*68*), which would be revolutionary performance improvement for TFTs processed in solution phase. TMDs have also shown encouraging performance in thin films, with mobilities of 1 to 40 $cm^2 V^{-1} s^{-1}$ (*96*–*100*), but they remain limited by insufficient synthesis to fabricate large numbers of transistors on a single substrate and the required use of high-temperature synthesis (not processed in solution). With cost being one of the most important driving factors for TFT applications, the most promising nanomaterial options are those that help break the cost barrier of current technologies. The most attractive way for this to happen is for a TFT to be printed rather processed in a vacuum or at a high temperature. This strength is beginning to receive attention for nanomaterial TFTs (*101*–*106*).

There are many other applications that benefit from current and future TFT technologies. Displays are presently the largest, and thus the most emphasized, but gas and biological sensors, optical detectors, radio-frequency identification tags, and Internet-of-Things applications also abound. Advancements in TFTs that allow for performance improvement and cost reduction are needed to open the way for a myriad of such exciting applications. Nanomaterials are poised to provide such solutions, requiring further improvement in synthesis and thin-film formation; discovery of better contact interfaces; and realization of stable, consistent TFT operation.

New generation of transistors

Thus far, the focus of this Review has been how nanomaterial transistors have the potential to provide transformative solutions in high-performance and thin-film applications. Yet there are other, completely new and distinct application spaces for which nanomaterial transistors are particularly suited. Just as TFT research is focused more on cost than performance, studies of these new generations of transistors focus on other opportunities, such as optical transparency or biocompatibility. A subset of new-generation transistor applications— including printed electronics (*103*), flexible electronics (*93*), transparent electronics (*40*), and biomedical electronics (*107*)—is depicted in Fig. 6. In addition to these areas (though not specifically highlighted here), nanomaterial transistors are beneficial in harsh environments [high temperature or radiation (*108*, *109*)] such as space or medical imaging applications, where the thinness and small cross-sectional area are among several nanomaterial advantages enabling greater protection from device damage.

The field of printed electronics has grown dramatically over the past 20 years. Relying almost exclusively on organic polymer materials, applications of printed electronics have been quite limited. Nanomaterials have the potential to offer conducting, semiconducting, and insulating printable inks that are compatible, stable, and able to be modulated for specific application needs. Such a toolkit of inks would be groundbreaking for this field that has potential to revolutionize on-the-fly electronics in an analogous fashion to how 3D printing of mechanical structures has transformed prototyping.

Nanomaterials offer some of the most promising possibilities for electronics that are flexible, transparent, or both. In addition to their substrate independence (discussed above), the atomic thinness of the nanomaterials also renders them nearly 100% transparent to visible light and has been exploited in some demonstrated all-2D transistors (*40*, *43*), motivated by the ability to create completely transparent circuits. Furthermore, the electronic properties of nanomaterials have proven to be robust to mechanical deformation for flexible substrate applications.

Finally, biomedical applications that have requirements of biocompatibility could also benefit from nanomaterials (*107*). For in vivo applications, small quantities of nanomaterials could be sufficiently safe to enable diagnostic circuits. Electronics applied directly to the skin are another possibility, where nanomaterials would mostly provide the types of benefits outlined in the flexible electronics discussion above.

Outlook

Transistors have completely transformed our daily lives, in areas ranging from communication to computation to comfort. As the limits of silicon transistors are unavoidably upon us, an intensified consideration of other transistor options is imperative. It is also important to keep clear the relevant deliverables for certain transistor applications spaces: namely, high performance and thin film. High-performance transistors are used in all computational devices, from servers to smartphones, whereas TFTs primarily provide the backplane circuitry for displays. The meaningfulness of parameters such as mobility must be kept in context for the different transistor applications so that scientific advances can be kept in proper perspective.

Nanomaterials, including carbon nanotubes and TMDs, show great promise for both high-performance transistors and TFTs. Not every scientific discovery or advancement benefits both nanomaterial high-performance transistors and TFTs, yet many reports confuse this fact. What is clear from the thousands of demonstrated nanomaterial transistors is that they offer considerable advantages, promoting transformative advancement in high-performance, thin-film, and completely new application spaces. Comparison of the distinctive aspects of the different nanomaterial options, or the transistors

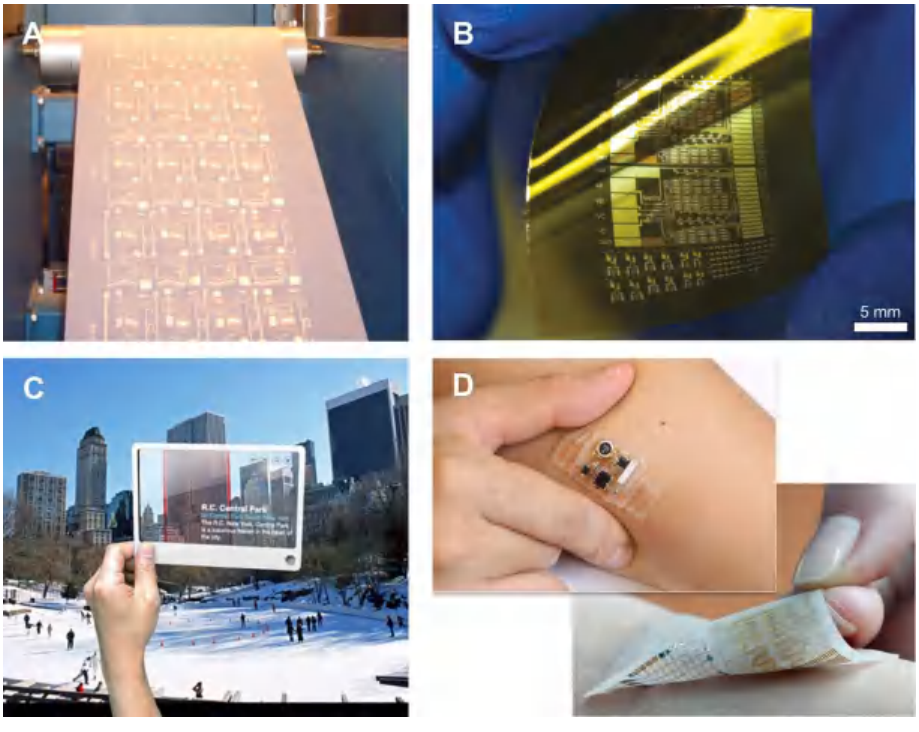


Fig. 6. New application spaces for nanomaterial transistors. (**A**) Printed electronics, where nanomaterials offer air-stable metallic, semiconducting, and even insulating options for the low-cost fabrication of multifunctional electronics. [From (*110*)] (**B**) Flexible electronics. Many demonstrations of nanomaterial transistors on flexible substrates have shown their robustness to mechanical deformation. [Modified from (*93*), with permission] (**C**) Transparent electronics can benefit greatly from the near complete optical transparency of nanomaterials. [From (*111*)] (**D**) Biomedical electronics require materials that can be safely dissolved on skin or even in vivo, and nanomaterials show promise for yielding the necessary function at quantities small enough to be safe. [Modified from (*107*), with permission]

assembled from them, suggests greater suitability of certain nanomaterials for particular applications (e.g., the instability of most X-enes in air largely precludes their use in low-cost, thin-film processing).

Going forward, much work remains in terms of improving the synthesis and resultant uniformity of nanomaterials, positioning them onto target device substrates, understanding and optimizing key interfaces (contacts and gate), and exploring re-invented transistor structures that make more direct use of the nanomaterial benefits. Varied levels of progress have been achieved in addressing these challenges for each nanomaterial, but each requires further attention to be suitable for a transistor technology. The most accessible area in the near term is the use of nanomaterials for printed (low-cost) TFTs, where they are already beginning to show consistent advantages over existing technologies.

Perhaps the greatest challenge to this unprecedentedly large research field of nanomaterial transistors is to make appropriate use of the deluge of scientific and engineering advancements. Hopefully, in the years to come, more collaborative and comprehensive advancements will be published that not only demonstrate a singular breakthrough on one of the challenges for nanomaterials but also give thorough evidence of the impacts of their solution on other key deliverables for a certain nanomaterial transistor application. With sufficient continued research investment into exploring nanomaterial transistors, there is little doubt that 50 years from now the next generation of scientists will be reflecting on a full century of Moore's law, in one form or another.

REFERENCES

1. G. E. Moore, Cramming more components onto integrated circuits. *Electronics* **38**, 114–117 (1965).
2. D. Rosso, "Global semiconductor industry posts record sales in 2014," Semiconductor Industry Association, press release (2 February 2015); www.semiconductors.org/news/2015/02/02/global_sales_report_2014/global_semiconductor_industry_posts_record_sales_in_2014/.
3. J. Hruska, "Stop obsessing over transistor counts: It's a terrible way of comparing chips," *ExtremeTech* (2014); www.extremetech.com/computing/190946-stop-obsessing-over-transistor-counts-theyre-a-terrible-way-of-comparing-chips.
4. R. A. Street, Thin-film transistors. *Adv. Mater.* **21**, 2007–2022 (2009). doi: 10.1002/adma.200803211; pmid: 21161982
5. Y. Kuo, Thin film transistor technology—Past, present, and future. *Interface* **22**, 55–61 (2013).
6. S. Bangsaruntip *et al*., "High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling," in *2009 IEEE International Electron Devices Meeting Technical Digest*, Baltimore, MD, 7 to 9 December 2009 (IEEE, 2009), pp. 297–300.
7. D. K. Ferry, Nanowires in nanoelectronics. *Science* **319**, 579–580 (2008). doi: 10.1126/science.1154446; pmid: 18239115
8. R. Ganatra, Q. Zhang, Few-layer MoS_2: A promising layered semiconductor. *ACS Nano* **8**, 4074–4099 (2014). doi: 10.1021/nn405938z; pmid: 24660756
9. A. B. Kaul, Two-dimensional layered materials: Structure, properties, and prospects for device applications. *J. Mater. Res.* **29**, 348–361 (2014). doi: 10.1557/jmr.2014.6
10. S. Z. Butler *et al*., Progress, challenges, and opportunities in two-dimensional materials beyond graphene. *ACS Nano* **7**, 2898–2926 (2013). doi: 10.1021/nn400280c; pmid: 23464873
11. M. C. Lemme, L.-J. Li, T. Palacios, F. Schwierz, Two-dimensional materials for electronic applications. *MRS Bull.* **39**, 711–718 (2014). doi: 10.1557/mrs.2014.138
12. M. Naguib, V. N. Mochalin, M. W. Barsoum, Y. Gogotsi, 25th anniversary article: MXenes: A new family of two-dimensional materials. *Adv. Mater.* **26**, 992–1005 (2014). doi: 10.1002/adma.201304138; pmid: 24357390
13. G. S. Tulevski *et al*., Toward high-performance digital logic technology with carbon nanotubes. *ACS Nano* **8**, 8730–8745 (2014). doi: 10.1021/nn503627h; pmid: 25144443
14. F. Schwierz, Graphene transistors. *Nat. Nanotechnol.* **5**, 487–496 (2010). doi: 10.1038/nnano.2010.89; pmid: 20512128
15. D. Reddy, L. F. Register, G. D. Carpenter, S. K. Banerjee, Graphene field-effect transistors. *J. Phys. D Appl. Phys.* **44**, 313001 (2011). doi: 10.1088/0022-3727/44/31/313001
16. Y. Zhang *et al*., Direct observation of a widely tunable bandgap in bilayer graphene. *Nature* **459**, 820–823 (2009). doi: 10.1038/nature08105; pmid: 19516337
17. F. Xia, D. B. Farmer, Y. M. Lin, P. Avouris, Graphene field-effect transistors with high on/off current ratio and large transport band gap at room temperature. *Nano Lett.* **10**, 715–718 (2010). doi: 10.1021/nl9039636; pmid: 20092332
18. X. Wang *et al*., Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors. *Phys. Rev. Lett.* **100**, 206803 (2008). doi: 10.1103/PhysRevLett.100.206803; pmid: 18518566
19. X. Liang, S. Wi, Transport characteristics of multichannel transistors made from densely aligned sub-10 nm half-pitch graphene nanoribbons. *ACS Nano* **6**, 9700–9710 (2012). doi: 10.1021/nn303127y; pmid: 23078122
20. S. Balendhran, S. Walia, H. Nili, S. Sriram, M. Bhaskaran, Elemental analogues of graphene: Silicene, germanene, stanene, and phosphorene. *Small* **11**, 640–652 (2015). doi: 10.1002/smll.201402041; pmid: 25380184
21. L. Tao *et al*., Silicene field-effect transistors operating at room temperature. *Nat. Nanotechnol.* **10**, 227–231 (2015). doi: 10.1038/nnano.2014.325; pmid: 25643256
22. C. Kamal, M. Ezawa, Arsenene: Two-dimensional buckled and puckered honeycomb arsenic systems. *Phys. Rev. B* **91**, 085423 (2015). doi: 10.1103/PhysRevB.91.085423
23. H. Liu *et al*., Phosphorene: An unexplored 2D semiconductor with a high hole mobility. *ACS Nano* **8**, 4033–4041 (2014). doi: 10.1021/nn501226z; pmid: 24655084
24. M. Derivaz *et al*., Continuous germanene layer on Al(111). *Nano Lett.* **15**, 2510–2516 (2015). doi: 10.1021/acs.nanolett.5b00085; pmid: 25802988
25. K. Lam, Z. Dong, J. Guo, Performance limits projection of black phosphorous field-effect transistors. *IEEE Electron Device Lett.* **35**, 963–965 (2014). doi: 10.1109/LED.2014.2333368
26. H. Liu, A. T. Neal, M. Si, Y. Du, P. D. Ye, The effect of dielectric capping on few-layer phosphorene transistors: Tuning the Schottky barrier heights. *IEEE Trans. Electron. Dev.* **35**, 795–797 (2014). doi: 10.1109/LED.2014.2323951
27. C. P. Auth, J. D. Plummer, Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's. *IEEE Electron Device Lett.* **18**, 74–76 (1997). doi: 10.1109/55.553049
28. B. S. Flavel, K. E. Moore, M. Pfohl, M. M. Kappes, F. Hennrich, Separation of single-walled carbon nanotubes with a gel permeation chromatography system. *ACS Nano* **8**, 1817–1826 (2014). doi: 10.1021/nn4062116; pmid: 24460395
29. M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, M. C. Hersam, Sorting carbon nanotubes by electronic structure using density differentiation. *Nat. Nanotechnol.* **1**, 60–65 (2006). doi: 10.1038/nnano.2006.52; pmid: 18654143
30. X. Tu, S. Manohar, A. Jagota, M. Zheng, DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes. *Nature* **460**, 250–253 (2009). doi: 10.1038/nature08116; pmid: 19587767
31. J. Liu, M. Hersam, Recent developments in carbon nanotube sorting and selective growth. *MRS Bull.* **35**, 315–321 (2010). doi: 10.1557/mrs2010.554
32. C. Y. Khripin, J. A. Fagan, M. Zheng, Spontaneous partition of carbon nanotubes in polymer-modified aqueous phases. *J. Am. Chem. Soc.* **135**, 6822–6825 (2013). doi: 10.1021/ja402762e; pmid: 23611526
33. G. S. Tulevski, A. D. Franklin, A. Afzali, High purity isolation and quantification of semiconducting carbon nanotubes via column chromatography. *ACS Nano* **7**, 2971–2976 (2013). doi: 10.1021/nn400053k; pmid: 23484490
34. C. Homenick *et al*., High-yield, single-step separation of metallic and semiconducting SWCNTs using block copolymers at low temperatures. *J. Phys. Chem. C* **118**, 16156–16164 (2014). doi: 10.1021/jp5030476
35. J. Park, P. Deria, J.-H. Olivier, M. J. Therien, Fluence-dependent singlet exciton dynamics in length-sorted chirality-enriched single-walled carbon nanotubes. *Nano Lett.* **14**, 504–511 (2014). doi: 10.1021/nl403511s; pmid: 24329134
36. K. S. Mistry, B. A. Larsen, J. L. Blackburn, High-yield dispersions of large-diameter semiconducting single-walled carbon nanotubes with tunable narrow chirality distributions. *ACS Nano* **7**, 2231–2239 (2013). doi: 10.1021/nn305336x; pmid: 23379962
37. S. Fogden, C. A. Howard, R. K. Heenan, N. T. Skipper, M. S. P. Shaffer, Scalable method for the reductive dissolution, purification, and separation of single-walled carbon nanotubes. *ACS Nano* **6**, 54–62 (2012). doi: 10.1021/nn2041494; pmid: 22206484
38. T. Tanaka, Y. Urabe, D. Nishide, H. Kataura, Discovery of surfactants for metal/semiconductor separation of single-wall carbon nanotubes via high-throughput screening. *J. Am. Chem. Soc.* **133**, 17610–17613 (2011). doi: 10.1021/ja208221g; pmid: 21980975
39. T. Roy *et al*., Field-effect transistors built from all two-dimensional material components. *ACS Nano* **8**, 6259–6264 (2014). doi: 10.1021/nn501723y; pmid: 24779528
40. S. Das, R. Gulotty, A. V. Sumant, A. Roelofs, All two-dimensional, flexible, transparent, and thinnest thin film transistor. *Nano Lett.* **14**, 2861–2866 (2014). doi: 10.1021/nl5009037; pmid: 24754722
41. A. Avsar *et al*., Air-stable transport in graphene-contacted, fully encapsulated ultrathin black phosphorus-based field-effect transistors. *ACS Nano* **9**, 4138–4145 (2015). doi: 10.1021/acsnano.5b00289; pmid: 25769342
42. K. G. Zhou *et al*., Raman modes of MoS_2 used as fingerprint of van der Waals interactions in 2-D crystal-based heterostructures. *ACS Nano* **8**, 9914–9924 (2014). doi: 10.1021/nn5042703; pmid: 25198732
43. G. H. Lee *et al*., Flexible and transparent MoS_2 field-effect transistors on hexagonal boron nitride-graphene heterostructures. *ACS Nano* **7**, 7931–7936 (2013). doi: 10.1021/nn402954e; pmid: 23924287
44. C. J. Shih *et al*., Tuning on-off current ratio and field-effect mobility in a MoS_2-graphene heterostructure via Schottky barrier modulation. *ACS Nano* **8**, 5790–5798 (2014). doi: 10.1021/nn500676t; pmid: 24824139
45. C. Huang *et al*., Lateral heterojunctions within monolayer $MoSe_2$-WSe_2 semiconductors. *Nat. Mater.* **13**, 1096–1101 (2014). doi: 10.1038/nmat4064; pmid: 25150560
46. H. Wang *et al*., Two-dimensional heterostructures: Fabrication, characterization, and application. *Nanoscale* **6**, 12250–12272 (2014). doi: 10.1039/C4NR03435J; pmid: 25219598
47. L. Yu *et al*., Graphene/MoS_2 hybrid technology for large-scale two-dimensional electronics. *Nano Lett.* **14**, 3055–3063 (2014). doi: 10.1021/nl404795z; pmid: 24810658
48. X. Hong *et al*., Ultrafast charge transfer in atomically thin MoS_2/WS_2 heterostructures. *Nat. Nanotechnol.* **9**, 682–686 (2014). doi: 10.1038/nnano.2014.167; pmid: 25150718
49. J. H. Yu *et al*., Vertical heterostructure of two-dimensional MoS_2 and WSe_2 with vertically aligned layers. *Nano Lett.* **15**, 1031–1035 (2015). doi: 10.1021/nl503897h; pmid: 25590995
50. J. Knoch, J. Appenzeller, Tunneling phenomena in carbon nanotube field-effect transistors. *Phys. Status Solidi A* **205**, 679–694 (2008). doi: 10.1002/pssa.200723528
51. A. D. Franklin, Electronics: The road to carbon nanotube transistors. *Nature* **498**, 443–444 (2013). doi: 10.1038/498443a; pmid: 23803839
52. J. Jeon *et al*., Controlling grain size and continuous layer growth in two-dimensional MoS_2 films for nanoelectronic device application. *IEEE Trans. NanoTechnol.* **14**, 238–242 (2015). doi: 10.1109/TNANO.2014.2381667
53. D. Dumcenco *et al*., Large-area epitaxial monolayer MoS_2. *ACS Nano* **9**, 4611–4620 (2015). doi: 10.1021/acsnano.5b01281; pmid: 25843548
54. C.-C. Huang *et al*., Scalable high-mobility MoS_2 thin films fabricated by an atmospheric pressure chemical vapor deposition process at ambient temperature. *Nanoscale* **6**, 12792–12797 (2014). doi: 10.1039/C4NR04228J; pmid: 25226424
55. Z. Jin, S. Shin, H. Kwon, S.-J. Han, Y.-S. Min, Novel chemical route for atomic layer deposition of MoS_2 thin film on SiO_2/Si substrate. *Nanoscale* **6**, 14453–14458 (2014). doi: 10.1039/C4NR04816D; pmid: 25340905
56. A. Tarasov *et al*., Highly uniform trilayer molybdenum disulfide for wafer-scale device fabrication. *Adv. Funct. Mater.* **24**, 6389–6400 (2014). doi: 10.1002/adfm.201401389

57. H. Schmidt *et al*., Transport properties of monolayer MoS_2 grown by chemical vapor deposition. *Nano Lett.* **14**, 1909–1913 (2014). doi: 10.1021/nl4046922; pmid: 24640984
58. M. Okada *et al*., Direct chemical vapor deposition growth of WS2 atomic layers on hexagonal boron nitride. *ACS Nano* **8**, 8273–8277 (2014). doi: 10.1021/nn503093k; pmid: 25093606
59. A. S. George *et al*., wafer scale synthesis and high resolution structural characterization of atomically thin MoS_2 layers. *Adv. Funct. Mater.* **24**, 7461–7466 (2014). doi: 10.1002/adfm.201402519
60. Y. Yu *et al*., Controlled scalable synthesis of uniform, high-quality monolayer and few-layer MoS_2 films. *Sci. Rep.* **3**, 1866 (2013). doi: 10.1038/srep01866; pmid: 23689610
61. S. Park *et al*., Large-area assembly of densely aligned single-walled carbon nanotubes using solution shearing and their application to field-effect transistors. *Adv. Mater.* **27**, 2656–2662 (2015). doi: 10.1002/adma.201405289; pmid: 25788393
62. Q. Cao *et al*., Arrays of single-walled carbon nanotubes with full surface coverage for high-performance electronics. *Nat. Nanotechnol.* **8**, 180–186 (2013). doi: 10.1038/nnano.2012.257; pmid: 23353673
63. J. M. Lobez, S.-J. Han, A. Afzali, J. B. Hannon, Surface selective one-step fabrication of carbon nanotube thin films with high density. *ACS Nano* **8**, 4954–4960 (2014). doi: 10.1021/nn5009935; pmid: 24684374
64. T. A. Shastry *et al*., Large-area, electronically monodisperse, aligned single-walled carbon nanotube thin films fabricated by evaporation-driven self-assembly. *Small* **9**, 45–51 (2013). doi: 10.1002/smll.201201398; pmid: 22987547
65. S.-P. Han, H. T. Maune, R. D. Barish, M. Bockrath, W. A. Goddard III, DNA-linker-induced surface assembly of ultra dense parallel single walled carbon nanotube arrays. *Nano Lett.* **12**, 1129–1135 (2012). doi: 10.1021/nl201818u; pmid: 22320204
66. S. Shekhar, P. Stokes, S. I. Khondaker, Ultrahigh density alignment of carbon nanotube arrays by dielectrophoresis. *ACS Nano* **5**, 1739–1746 (2011). doi: 10.1021/nn102305z; pmid: 21323326
67. H. Park *et al*., High-density integration of carbon nanotubes via chemical self-assembly. *Nat. Nanotechnol.* **7**, 787–791 (2012). doi: 10.1038/nnano.2012.189; pmid: 23103933
68. G. J. Brady *et al*., Polyfluorene-sorted, carbon nanotube array field-effect transistors with increased current density and high on/off ratio. *ACS Nano* **8**, 11614–11621 (2014). doi: 10.1021/nn5048734; pmid: 25383880
69. M. Shimizu, S. Fujii, T. Tanaka, H. Kataura, Effects of surfactants on the electronic transport properties of thin-film transistors of single-wall carbon nanotubes. *J. Phys. Chem. C* **117**, 11744–11749 (2013). doi: 10.1021/jp3113254
70. V. K. Sangwan *et al*., Fundamental performance limits of carbon nanotube thin-film transistors achieved using hybrid molecular dielectrics. *ACS Nano* **6**, 7480–7488 (2012). doi: 10.1021/nn302768h; pmid: 22783918
71. S.-J. Choi *et al*., Comparative study of solution-processed carbon nanotube network transistors. *Appl. Phys. Lett.* **101**, 112104 (2012). doi: 10.1063/1.4752006
72. D. B. Farmer, R. G. Gordon, Atomic layer deposition on suspended single-walled carbon nanotubes via gas-phase noncovalent functionalization. *Nano Lett.* **6**, 699–703 (2006). doi: 10.1021/nl052453d; pmid: 16608267
73. A. D. Franklin *et al*., Carbon nanotube complementary wrap-gate transistors. *Nano Lett.* **13**, 2490–2495 (2013). doi: 10.1021/nl400544q; pmid: 23638708
74. S. Jandhyala *et al*., Atomic layer deposition of dielectrics on graphene using reversibly physisorbed ozone. *ACS Nano* **6**, 2722–2730 (2012). doi: 10.1021/nn300167t; pmid: 22352388
75. A. Azcatl *et al*., MoS_2 functionalization for ultra-thin atomic layer deposited dielectrics. *Appl. Phys. Lett.* **104**, 111601 (2014). doi: 10.1063/1.4869149
76. Z. Chen, J. Appenzeller, J. Knoch, Y. M. Lin, P. Avouris, The role of metal-nanotube contact in the performance of carbon nanotube field-effect transistors. *Nano Lett.* **5**, 1497–1502 (2005). doi: 10.1021/nl0508624; pmid: 16178264
77. Y. Chai *et al*., Low-resistance electrical contact to carbon nanotubes with graphitic interfacial layer. *IEEE Trans. Electron. Dev.* **59**, 12–19 (2012). doi: 10.1109/TED.2011.2170216
78. A. D. Franklin, D. B. Farmer, W. Haensch, Defining and overcoming the contact resistance challenge in scaled carbon nanotube transistors. *ACS Nano* **8**, 7333–7339 (2014). doi: 10.1021/nn5024363; pmid: 24999536
79. W. Liu *et al*., Role of metal contacts in designing high-performance monolayer n-type WSe2 field effect transistors. *Nano Lett.* **13**, 1983–1990 (2013). doi: 10.1021/nl304777e; pmid: 23527483
80. S. Das, H.-Y. Chen, A. V. Penumatcha, J. Appenzeller, High performance multilayer MoS_2 transistors with scandium contacts. *Nano Lett.* **13**, 100–105 (2013). doi: 10.1021/nl303583v; pmid: 23240655
81. C. D. English, G. Shine, V. E. Dorgan, K. C. Saraswat, E. Pop, Improving contact resistance in MoS_2 field effect transistors. *Device Res. Conf. Dig.* **2014**, 193–194 (2014).
82. A. Chanana, S. Mahapatra, First principles study of metal contacts to monolayer black phosphorous. *J. Appl. Phys.* **116**, 204302 (2014). doi: 10.1063/1.4901998
83. There is no longer much meaning behind this dimensional indicator for a certain technology node, as discussed in http://spectrum.ieee.org/semiconductors/devices/the-status-of-moores-law-its-complicated, where it is noted that the 14-nm node has 35-nm-long channel lengths with 8-nm-wide silicon fins.
84. E. Pop, Energy dissipation and transport in nanoscale devices. *Nano Res.* **3**, 147–169 (2010). doi: 10.1007/s12274-010-1019-z
85. H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, D. Burger, "Dark silicon and the end of multicore scaling," in *ISCA '11 Proceedings of the 38th Annual International Symposium on Computer Architecture*, San Jose, CA, 4 to 8 June 2011 (Association for Computing Machinery, New York, 2011), pp. 365–376.
86. J. Appenzeller, Y. Lin, J. Knoch, Z. Chen, P. Avouris, comparing carbon nanotube transistors - the ideal choice: A novel tunneling device design. *IEEE Trans. Electron. Dev.* **52**, 2568–2576 (2005). doi: 10.1109/TED.2005.859654
87. D. Jena, Tunneling transistors based on graphene and 2-D crystals. *Proc. IEEE* **101**, 1585–1602 (2013). doi: 10.1109/JPROC.2013.2253435
88. A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai, Ballistic carbon nanotube field-effect transistors. *Nature* **424**, 654–657 (2003). doi: 10.1038/nature01797; pmid: 12904787
89. A. D. Franklin, Z. Chen, Length scaling of carbon nanotube transistors. *Nat. Nanotechnol.* **5**, 858–862 (2010). doi: 10.1038/nnano.2010.220; pmid: 21102468
90. A. D. Franklin *et al*., Sub-10 nm carbon nanotube transistor. *Nano Lett.* **12**, 758–762 (2012). doi: 10.1021/nl203701g; pmid: 22260387
91. S. M. Sze, *Physics of Semiconductor Devices* (Wiley, Hoboken, NJ, ed. 2, 1981).
92. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge Univ. Press, New York, 2005).
93. Q. Cao *et al*., Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates. *Nature* **454**, 495–500 (2008). doi: 10.1038/nature07110; pmid: 18650920
94. B. Chandra, H. Park, A. Maarouf, G. J. Martyna, G. S. Tulevski, Carbon nanotube thin film transistors on flexible substrates. *Appl. Phys. Lett.* **99**, 072110 (2011). doi: 10.1063/1.3622767
95. D. Lee *et al*., High-performance thin-film transistors produced from highly separated solution-processed carbon nanotubes. *Appl. Phys. Lett.* **104**, 143508 (2014). doi: 10.1063/1.4871100
96. G. A. Salvatore *et al*., Fabrication and transfer of flexible few-layers MoS_2 thin film transistors to any arbitrary substrate. *ACS Nano* **7**, 8809–8815 (2013). doi: 10.1021/nn403248y; pmid: 23991756
97. H.-Y. Chang *et al*., High-performance, highly bendable MoS_2 transistors with high-k dielectrics for flexible low-power systems. *ACS Nano* **7**, 5446–5452 (2013). doi: 10.1021/nn401429w; pmid: 23668386
98. R. Yang, Z. Wang, P. X.-L. Feng, Electrical breakdown of multilayer MoS_2 field-effect transistors with thickness-dependent mobility. *Nanoscale* **6**, 12383–12390 (2014). doi: 10.1039/C4NR03472D; pmid: 25219778
99. N. R. Pradhan *et al*., Field-effect transistors based on few-layered α-$MoTe_2$. *ACS Nano* **8**, 5911–5920 (2014). doi: 10.1021/nn501013c; pmid: 24878323
100. J. Kumar, M. A. Kuroda, M. Z. Bellus, S.-J. Han, H.-Y. Chiu, Full-range electrical characteristics of WS_2 transistors. *Appl. Phys. Lett.* **106**, 123508 (2015). doi: 10.1063/1.4916403
101. B. Kim *et al*., High-speed, inkjet-printed carbon nanotube/zinc tin oxide hybrid complementary ring oscillators. *Nano Lett.* **14**, 3683–3687 (2014). doi: 10.1021/nl5016014; pmid: 24849313
102. M. Ha *et al*., Aerosol jet printed, low voltage, electrolyte gated carbon nanotube ring oscillators with sub-5 μs stage delays. *Nano Lett.* **13**, 954–960 (2013). doi: 10.1021/nl3038773; pmid: 23394463
103. P. Chen *et al*., Fully printed separated carbon nanotube thin film transistor circuits and its application in organic light emitting diode control. *Nano Lett.* **11**, 5301–5308 (2011). doi: 10.1021/nl202765b; pmid: 22050730
104. N. Rouhi, D. Jain, P. J. Burke, High-performance semiconducting nanotube inks: Progress and prospects. *ACS Nano* **5**, 8471–8487 (2011). doi: 10.1021/nn201828y; pmid: 21970293
105. F. Torrisi, J. N. Coleman, Electrifying inks with 2D materials. *Nat. Nanotechnol.* **9**, 738–739 (2014). doi: 10.1038/nnano.2014.218; pmid: 25286261
106. F. Torrisi *et al*., Inkjet-printed graphene electronics. *ACS Nano* **6**, 2992–3006 (2012). doi: 10.1021/nn2044609; pmid: 22449258
107. D. Son *et al*., Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nat. Nanotechnol.* **9**, 397–404 (2014). doi: 10.1038/nnano.2014.38; pmid: 24681776
108. C. D. Cress *et al*., Radiation effects in carbon nanoelectronics. *Electronics* **1**, 23–31 (2012). doi: 10.3390/electronics1010023
109. O. Ochedowski *et al*., Radiation hardness of graphene and MoS_2 field effect devices against swift heavy ion irradiation. *J. Appl. Phys.* **113**, 214306 (2013). doi: 10.1063/1.4808460
110. IDTechEx, "End user needs and case studies aired at IDTechEx's Printed Electronics Europe," *Print. Electron. News*, press release (28 March 2011).
111. M. Wilson, "Researchers develop transparent memory, see-through electronics next," (2008); http://gizmodo.com/5111411/researchers-develop-transparent-memory-see-through-electronics-next.

10.1126/science.aab2750

RESEARCH ARTICLE SUMMARY

HIV-1 VACCINES

Diversion of HIV-1 vaccine–induced immunity by gp41-microbiota cross-reactive antibodies

Wilton B. Williams,* Hua-Xin Liao, M. Anthony Moody, Thomas B. Kepler, S. Munir Alam, Feng Gao, Kevin Wiehe, Ashley M. Trama, Kathryn Jones, Ruijun Zhang, Hongshuo Song, Dawn J. Marshall, John F. Whitesides, Kaitlin Sawatzki, Axin Hua, Pinghuang Liu, Matthew Z. Tay, Kelly E. Seaton, Xiaoying Shen, Andrew Foulger, Krissey E. Lloyd, Robert Parks, Justin Pollara, Guido Ferrari, Jae-Sung Yu, Nathan Vandergrift, David C. Montefiori, Magdalena E. Sobieszczyk, Scott Hammer, Shelly Karuna, Peter Gilbert, Doug Grove, Nicole Grunenberg, M. Juliana McElrath, John R. Mascola, Richard A. Koup, Lawrence Corey, Gary J. Nabel,† Cecilia Morgan, Gavin Churchyard, Janine Maenza, Michael Keefer, Barney S. Graham, Lindsey R. Baden, Georgia D. Tomaras, Barton F. Haynes*

INTRODUCTION: Inducing protective antibodies is a key goal in HIV-1 vaccine development. In acute HIV-1 infection, the dominant initial plasma antibody response is to the gp41 subunit of the envelope (Env) glycoprotein of the virus. These antibodies derive from polyreactive B cells that cross-react with Env and intestinal microbiota (IM) and are unable to neutralize HIV-1. However, whether a similar gp41-IM cross-reactive antibody response would occur in the setting of HIV-1 Env vaccination is unknown.

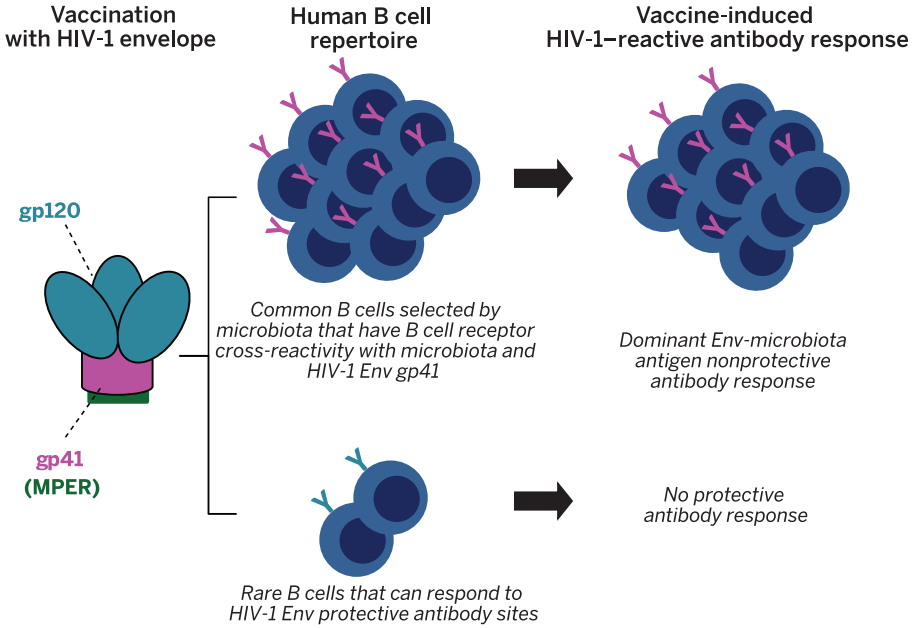
RATIONALE: We studied antibody responses in individuals who received a DNA prime vaccine, with a recombinant adenovirus serotype 5 (rAd5) boost (DNA prime–rAd5 boost), a vaccine that included HIV-1 *gag*, *pol*, and *nef* genes, as well as a trivalent mixture of clade A, B, and C *env* gp140 genes containing both gp120 and gp41 components. This vaccine showed no efficacy. Thus, study of these vaccinees provided an opportunity to determine whether the Env-reactive antibody response in the setting of Env vaccination was dominated by gp41-reactive antibodies derived from Env-IM cross-reactive B cells.

RESULTS: We found that vaccine-induced antibodies to HIV-1 Env dominantly focused on gp41 compared with gp120 by both serologic analysis and by vaccine-Env memory B cells sorted by flow cytometry (see the figure). Remarkably, the majority of HIV-1 Env-reactive memory B cells induced by the vaccine produced gp41-reactive antibodies, and the majority of gp41-targeted antibodies used restricted immunoglobulin heavy chain variable genes. Functionally, none of the gp41-reactive antibodies could neutralize HIV, and the majority could not mediate antibody-dependent cellular cytotoxicity. Most of the vaccine-induced gp41-reactive antibodies cross-reacted with host and IM antigens. Two of the candidate gp41-intestinal cross-reactive antigens were bacterial RNA polymerase and pyruvate-flavodoxin oxidoreductase, which shared sequence similarities with the heptad repeat 1 region of HIV gp41. Next-generation sequencing of vaccinee B cells demonstrated a prevaccination antibody that was reactive to both IM and the vaccine–Env gp140, which demonstrated the presence of a preexisting pool of gp41-IM cross-reactive B cells from which the vaccine gp41-reactive antibody response was derived.

CONCLUSION: In this study, we found that the DNA prime–rAd5 boost HIV-1 vaccine induced a gp41-reactive antibody response that was mainly non-neutralizing and derived from an IM-gp41 cross-reactive B cell pool. These findings have important implications for HIV-1 vaccine design. Because IM antigens shape the B cell repertoire from birth, our data raise the hypothesis that neonatal immunization with HIV-1 envelope may be able to imprint the B cell repertoire to respond to envelope antigenic sites that may otherwise be subdominant or disfavored, such as Env broadly neutralizing antibody epitopes. Our data also suggest that deleting or modifying amino acids in the gp41 heptad repeat 1 region of Env-containing vaccine immunogens may avoid IM-gp41 cross-reactivity. Thus, an obstacle that may need to be overcome for development of a successful HIV vaccine is diversion of potentially protective HIV-1 antibody responses by preexisting envelope-IM cross-reactive pools of B cells. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aab1253



Diversion of HIV-1 vaccine–induced immunity by Env gp41–microbiota cross-reactive antibodies. Immunization of humans with a vaccine containing HIV-1 Env gp120 and gp41 components, including the membrane-proximal external region (MPER) of Env, induced a dominant B cell response primarily from a preexisting pool of gp41-IM cross-reactive B cells. This response diverted the vaccine-stimulated antibody response away from smaller subdominant B cell pools capable of reacting with potentially protective epitopes on HIV-1 Env.

The list of author affiliations is available in the full article online.
†Present address: Sanofi, 640 Memorial Drive, Cambridge, MA 02139, USA.
*Corresponding author. E-mail: barton.haynes@duke.edu (B.F.H.); wilton.williams@duke.edu (W.B.W.)
Cite this article as W. B. Williams *et al.*, *Science* **349**, aab1253 (2015). DOI: 10.1126/science.aab1253

RESEARCH ARTICLE

HIV-1 VACCINES

Diversion of HIV-1 vaccine–induced immunity by gp41-microbiota cross-reactive antibodies

Wilton B. Williams,[1]* Hua-Xin Liao,[1] M. Anthony Moody,[1] Thomas B. Kepler,[2] S. Munir Alam,[1] Feng Gao,[1] Kevin Wiehe,[1] Ashley M. Trama,[1] Kathryn Jones,[1] Ruijun Zhang,[1] Hongshuo Song,[1] Dawn J. Marshall,[1] John F. Whitesides,[1] Kaitlin Sawatzki,[2] Axin Hua,[2] Pinghuang Liu,[1] Matthew Z. Tay,[1] Kelly E. Seaton,[1] Xiaoying Shen,[1] Andrew Foulger,[1] Krissey E. Lloyd,[1] Robert Parks,[1] Justin Pollara,[1] Guido Ferrari,[1] Jae-Sung Yu,[1] Nathan Vandergrift,[1] David C. Montefiori,[1] Magdalena E. Sobieszczyk,[3] Scott Hammer,[3] Shelly Karuna,[4] Peter Gilbert,[5] Doug Grove,[5] Nicole Grunenberg,[4] M. Juliana McElrath,[4] John R. Mascola,[6] Richard A. Koup,[6] Lawrence Corey,[4] Gary J. Nabel,[6]† Cecilia Morgan,[5] Gavin Churchyard,[7] Janine Maenza,[4] Michael Keefer,[8] Barney S. Graham,[6] Lindsey R. Baden,[9] Georgia D. Tomaras,[1] Barton F. Haynes[1]*

An HIV-1 DNA prime vaccine, with a recombinant adenovirus type 5 (rAd5) boost, failed to protect from HIV-1 acquisition. We studied the nature of the vaccine-induced antibody (Ab) response to HIV-1 envelope (Env). HIV-1–reactive plasma Ab titers were higher to Env gp41 than to gp120, and repertoire analysis demonstrated that 93% of HIV-1–reactive Abs from memory B cells responded to Env gp41. Vaccine-induced gp41-reactive monoclonal antibodies were non-neutralizing and frequently polyreactive with host and environmental antigens, including intestinal microbiota (IM). Next-generation sequencing of an immunoglobulin heavy chain variable region repertoire before vaccination revealed an Env-IM cross-reactive Ab that was clonally related to a subsequent vaccine-induced gp41-reactive Ab. Thus, HIV-1 Env DNA-rAd5 vaccine induced a dominant IM-polyreactive, non-neutralizing gp41-reactive Ab repertoire response that was associated with no vaccine efficacy.

In acute HIV-1 infection, the dominant initial plasma antibody (Ab) response is to the gp41 subunit of the envelope (Env) glycoprotein of the virus (*1*). This Ab response derives from polyreactive B cells that cross-react with Env and intestinal microbiota (IM) (*2*, *3*). However, it is unknown if a similar gp41-reactive Ab response would occur in the setting of HIV-1 Env vaccination. A DNA prime vaccine, with a recombinant adenovirus serotype 5 (rAd5) boost (DNA prime–rAd5 boost), a vaccine that included HIV *gag*, *pol*, and *nef* genes—as well as a trivalent mixture of clade A, B, and C *env* gp140 genes containing both gp120 and gp41 components—was studied in the HIV Vaccine Trials Network (HVTN) [phase Ib (HVTN 082), phase II (HVTN 204), and phase IIb (HVTN 505) efficacy trial] and other clinical trials [phase I/II (RV172) and phase I (V001)] (*4*–*7*). This vaccine was the first vaccine containing the ectodomain of the Env gp41 component, covalently linked to gp120, to be tested in an efficacy trial and was designed to generate primarily CD8 T cell responses, although this vaccine generated Env Ab responses as well (*8*–*10*). However, the phase IIb HVTN 505 efficacy trial showed no vaccine efficacy (*11*). Thus, these vaccine trials containing Env gp41 provided an opportunity to determine whether the Env Ab response in the setting of Env vaccination was dominated by gp41-reactive Abs derived from Env-IM cross-reactive B cells.

[1]Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC, USA. [2]Department of Microbiology, Boston University School of Medicine, Boston, MA, USA. [3]Department of Medicine, Columbia University Medical Center, New York, NY, USA. [4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [5]The Statistical Center for HIV/AIDS Research and Prevention (SCHARP), Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [6]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [7]The Aurum Institute, Johannesburg, South Africa. [8]University of Rochester School of Medicine, Rochester, NY, USA. [9]Brigham and Women's Hospital, Boston, MA, USA.
*Corresponding author. E-mail: barton.haynes@duke.edu (B.F.H.); wilton.williams@duke.edu (W.B.W.) †Present address: Sanofi, 640 Memorial Drive, Cambridge, MA 02139, USA.

Isolation of Env-reactive memory B cells and vaccinee plasma serologies

We found that the DNA prime–rAd5 boost Ab response to HIV-1 Env was dominantly focused on gp41 compared with gp120. This specificity was demonstrated by both serologic analysis and vaccine-Env flow cytometry–sorted memory B cells. Plasma immunoglobulin G (IgG)–binding assays were performed on plasma of a random sample of 40 phase IIb (efficacy trial) vaccine recipients who were HIV-1 negative at the final, month 24, visit (*11*) (Fig. 1A), as well as plasma of eight HIV-1 uninfected phase Ib and II DNA prime–rAd5 boost trial participants with high titers of plasma-binding Abs to recombinant (r) gp140 vaccine–Envs and/or neutralization of clade C HIV-1 isolate MW965 (Fig. 1B). Plasma-binding gp41-reactive Ab titers were ≥10 times as much as gp120-reactive Ab titers, including Ab reactivity with vaccine-gp120s [($P < 0.0001$) (Fig. 1A), $P < 0.01$ (Fig. 1B); Wilcoxon signed rank test]. Thus, the nonprotective DNA prime–rAd5 boost gp140 vaccine induced a dominant HIV-1 Env gp41 response in plasma Abs.

Next, we performed single memory B cell sorting by flow cytometry using peripheral blood B cells from phase Ib and phase II DNA prime, rAd5-boost, trial participants. Vaccine-Env gp140 and V1V2 subunits—as well as a consensus group M gp140 Env (termed CON-S) (*12*)—were used as fluorophore-labeled recombinant proteins to identify Env-specific memory B cells present in peripheral blood mononuclear cells (PBMCs) of vaccinees 4 weeks after final vaccination (fig. S1 and table S1). We studied eight phase Ib and phase II DNA prime, rAd5-boost, trial participants; from these eight vaccinees, we isolated 221 HIV-1 Env-reactive Abs (Fig. 1C and table S2). Of the 221 HIV-1 Env-reactive Abs, there were 131 unique V_HDJ_H rearrangements (table S3). Remarkably, 205 out of 221 (205/221) (93%) of the HIV-1 Env-reactive Abs and 115/131 (88%) of the unique heavy chain sequences induced by the vaccine were gp41-reactive, with only 7% (16/221) gp120-reactive (tables S3 to S6). We used Ab gene transient transfections to perform enzyme-linked immunosorbent assays (ELISAs) to determine gp41 versus gp120 reactivity (*13*). Of the Env-reactive Abs, 16/16 (100%) gp120-reactive and 195/205 (95%) gp41-reactive Abs bound vaccine-rgp140 proteins. The 10 gp41-reactive Abs that bound only heterologous recombinant Env proteins likely recognized gp41 epitopes expressed on the vaccine protein generated by DNA or rAd5 that were not expressed on the rgp140 proteins.

We asked if there were indeed fewer gp120-reactive memory B cells from gp140-vaccinated individuals who received the DNA prime–rAd5 boost vaccine compared with gp140-reactive memory B cells. In three phase II trial vaccinee memory B cell samples, we found that Vaccine Research Center (VRC)–A gp120 bound to 0.37% of memory B cells compared with 0.55% of memory B cells by VRC-A gp140 ($P < 0.001$; Cochran-Mantel-Haenszel test) (fig. S2). Of the gp41-reactive and gp120-reactive Abs isolated from the eight phase Ib and II vaccinees, 112/221 (52%) were sorted using the group M consensus CON-S gp140 as a fluorophore-labeled hook. CON-S rgp120 bound to 0.05% memory B cells in comparison with 0.28% by CON-S gp140 ($P < 0.001$, Cochran-Mantel-Haenszel test). Of total HIV-1 Env-reactive memory B cells, CON-S gp120 bound to 13% (38/292), whereas 87% of Env-reactive memory B cells (254/292) reacted with CON-S gp140 (fig. S2). Therefore, the dearth of isolated gp120-reactive

Abs was mirrored by low frequencies of gp120-memory B cells in the blood of vaccinees.

For comparison, we have previously studied an Env gp120-only immunization trial in humans and demonstrated that at peak immunization the frequency of memory B cells ranged as high as 0.73% gp120-reactive B cells (mean 0.23 ± 0.1%) (*14*). In a rhesus macaque immunization study with gp120 alone, the frequency of memory B cells that were gp120-reactive was 3% (*15*). Thus, in the setting of gp120-only immunizations in humans and rhesus macaques, the memory B cell response to gp120 is robust.

Heavy chain gene restriction for HIV-1 gp41-reactive memory B cell Abs

We analyzed the variable heavy chain gene families used by HIV-1–reactive Abs induced by the DNA prime–rAd5 boost gp140 vaccine. Gp41-reactive Abs had a mean heavy chain nucleotide mutation frequency of 2.5% (fig. S3) and a median heavy chain complementarity-determining region 3 (HCDR3) length of 14 amino acids (fig. S4). Remarkably, of 137 total IGHV1-using Abs, 125 (91% of IGHV1-using Abs, 57% of total 221 HIV-1–reactive Abs) used a single V_H segment, IGHV1–69 (table S7). To rule out IGHV1-69 predominance as a result of polymerase chain reaction (PCR) primer bias, we examined the heavy chain gene usage of non–HIV-1–reactive Abs isolated from all eight vaccinees; only 12% (18/145) of non–HIV-1–reactive Abs used IGHV1-69 ($P < 0.0001$, Fisher's exact test). The 125 IGHV1-69–using HIV-1–reactive Abs all reacted with gp41, and 94 (75%) were found to be naturally paired with a kappa light chain and 31 (25%) with a lambda light chain. IGκV3-20 (48%, 60/94) and IGλ2-14 (29%, 9/31) were the light chains preferentially paired with the IGHV1-69–using gp41-reactive Abs (table S8). Of gp41-reactive Abs, 66% (136 of 205 Abs) used the *IGHV1* gene family; 61% (125 of 205 gp41-reactive Abs) specifically used the *IGHV1-69* gene segment (table S9), compared with HIV-1–uninfected individuals, in whom 6% of the Ab repertoire used IGHV1-69 (*16*) ($P < 0.0001$, chi-square test). Of the gp41-reactive Abs, 132 paired with IGκV light chains (table S10), whereas 73 gp41-reactive Abs paired with IGλV light chains (table S11). Moreover, next-generation sequencing (NGS) of prevaccination heavy chain V_HDJ_H rearrangements derived from RNA obtained from PBMCs of all eight vaccinees revealed ~5% IGHV1-69–using B cells before vaccination (tables S12). Thus, the frequency of IGHV1-69–using B cells was the same in healthy subjects as in prevaccination B cells but was dominantly used in the gp41 postvaccination repertoire, which indicated that the postvaccination gp41-reactive Ab IGHV1-69 usage induced by the DNA prime–rAd5 boost vaccine was not due to heavy chain gene primer bias but rather was selected by the vaccine.

There are 13 known allelic variants of *IGHV1-69* gene segment; 7 have a phenylalanine (F) at amino acid position 54 in the HCDR2 region (54F-HCDR2), and 6 have a leucine (L) (54L-HCDR2) (*17*). The ratio of 54F-HCDR2 to 54L-HCDR2 IGHV1-69 allelic variants in the global population is estimated to be 3:2 (*18*). Neutralizing influenza and HIV-1–reactive Abs use 54F-HCDR2 *IGHV1-69* gene segments to bind hydrophobic pockets in the stems of hemagglutinin (HA) (*19*, *20*) and gp41 (*21*, *22*), respectively. Thus, we asked whether both allelic forms of the *IGHV1-69* gene segment were equally represented in HIV-1 gp41-reactive Abs elicited by the DNA prime–rAd5 boost vaccine. Of 125 vaccine-induced IGHV1-69–using Abs, we found that 116 (93%) used 54L-HCDR2 variants, whereas only 9 (7%) used 54F-HCDR2 variants (Table 1 and table S5). NGS of prevaccination IGHV repertoire demonstrated that all eight phase Ib and II trial vaccinees encoded 54F- and 54L-HCDR2 variants of IGHV1-69 (table S12). Therefore, the expression of 54L-HCDR2 IGHV1-69 variants was not due to vaccinee inability to express 54F-HCDR2 variants. As a comparison, we asked what *IGHV1-69* gene allelic variants were used by gp41-reactive Abs in HIV-1 infection. Of 42 IGHV1-69–using gp41-reactive Abs from HIV-1–infected individuals (*23*), 41/42 (98%) used 54L-HCDR2 variants and 1/42 (2%) used a 54F-HCDR2 variant (Table 1). In contrast, of 64 non–HIV-1–reactive IGHV1-69–using Abs isolated from HIV-1 infected individuals (*2*, *3*, *23*), 18/64 (28%) used 54L-HCDR2 variants, whereas 46/64 (72%) used 54F-HCDR2 variants (Table 1). Finally, of 10 IGHV1-69–using influenza hemagglutinin Abs isolated from influenza vaccinees or acute influenza infection (*24*), 9 (90%) used 54F-HCDR2 variants and only 1 (10%) used a 54L-HCDR2 variant (Table 1). Thus, the DNA prime–rAd5 boost vaccine induced a dominant gp41-reactive Ab response that preferentially used 54L-HCDR2 IGHV1-69 allelic variants, similar to gp41-reactive Abs induced by HIV-1 infection ($P = 0.45$; Fisher's exact test) but different from influenza-induced IGHV1-69 Abs ($P < 0.0001$; Fisher's exact test) (Table 1 and table S5).

Because a subset of B-chronic lymphocytic leukemia (B-CLL) Abs uses IGHV1-69 and can

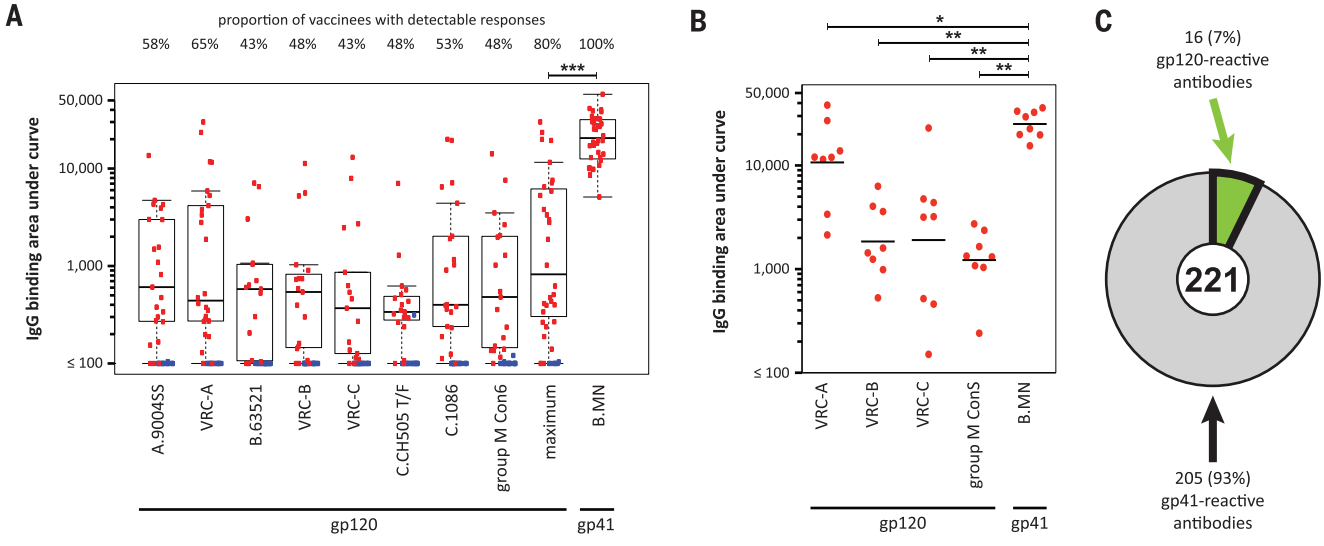


Fig. 1. Characteristics of HIV-1–reactive Abs induced by DNA prime–rAd5 boost vaccine. (**A**) Plasma Ab titers (AUC, area under the curve) to gp120 and gp41 proteins in a subset of 40 HVTN 505 vaccine recipients; red circles represent vaccine responders (percentages) to the antigens tested, and blue represent nonresponders. (**B**) Plasma binding Abs from eight HVTN 082 and 204 trial participants (vaccine responders) were screened for gp120 and gp41 reactivity by means of binding Ab multiplex assay (BAMA). (**C**) HIV-1–reactive Abs were isolated from single memory B cells of the eight HVTN 082 and 204 trial participants and screened for binding gp120 and gp41 proteins by means of enzyme-linked immunosorbent assay (ELISA). Statistics: (A) ***$P < 0.0001$, Wilcoxon signed rank test, B.MN gp41 versus gp120 (maximum; which is the highest value of all gp120 Envs for each participant), *$P < 0.05$, McNemar's test; B.MN gp41 versus each gp120 (not shown); (B) *$P < 0.05$, **$P < 0.01$, Wilcoxon signed rank test [Graphpad Prism].

Table 1. Frequency of 54L- and 54F-HCDR2–bearing IGHV1-69–using Abs induced by HIV and influenza. Counts reflect number of unique Ab heavy chain sequences containing full V_HDJ_H rearrangements without stop codons. Cohorts: HIV-1 vaccination[1], 8 HVTN 204 and 082 trial participants; HIV-1 infection[2,3], 29 HIV-1–infected subjects (*2*, *3*, *23*); influenza[4], 13 influenza-infected or vaccinated subjects (*24*). Influenza Ab specificity as measured by reactivity with purified recombinant HA-enriched trivalent influenza vaccine (Fluzone)[5]. Statistics were produced by using Fisher's exact test and SAS.

Cohort	Ab reactivity	Total Abs	IGHV1-69 Abs	HCDR2-54L	HCDR2-54F	Statistics HCDR2-54L:54F
HIV-1 vaccination[1]	gp41	205	125/205 (61%)	116/125 (93%)	9/125 (7%)	$P < 0.45^{(1v2)}$
HIV-1 infection[2]	gp41	116	42/116 (36%)	41/42 (98%)	1/42 (2%)	
HIV-1 infection[3]	Non–HIV-1	971	64/971 (7%)	18/64 (28%)	46/64 (72%)	$P < 0.0001^{(1v3,\ 2v3)}$
Influenza infection or vaccination[4]	Hemagglutinin[5]	278	10/278 (4%)	1/10 (10%)	9/10 (90%)	$P < 0.0001^{(1v4\ 2v4)}$; $P = 0.44^{(3v4)}$

cross-react with HIV-1 Env gp41 (*23*), we compared IGHV1-69 sequences of healthy control Abs (*16*), B-CLL Abs (*23*), and gp41-reactive Abs from phase Ib and II trials vaccine-recipients, for CLL archetypes previously described (*25*). We found that the frequency of CLL archetype matches in our vaccine-induced IGHV1-69–using Abs (1.6%) was no different from healthy control IGHV1-69 sequences (2.3%) but differed from CLL Abs (10.6%) (vaccine-induced gp41-reactive versus CLL Abs, $P = 0.002$; vaccine-induced gp41-reactive Abs versus healthy controls, NS; CLL versus healthy controls 1.7×10^{-5}, Fisher's exact test) (table S13). Thus, there is no evidence of selective derivation of vaccine-induced gp41-reactive IGHV1-69 Abs from the same pool of B cells that give rise to CLL Abs.

Functional properties of HIV-1–reactive mAbs

We chose 29 naturally paired heavy and light chain genes for recombinant Ab bulk expression and characterization; 17 gp41-reactive [representing 45 Abs within 17 vaccine-induced clonal lineages (table S14)] and 12 gp120-reactive [representing 12 clonal lineages (table S15)] monoclonal antibodies (mAbs). Gp41-reactive mAbs bound to rgp140 and/or gp41 proteins via ELISA and/or surface plasmon resonance (SPR) (table S16). For Abs that reacted strongly to vaccine Envs, the dissociation constants (K_d) for binding to gp41 MN recombinant protein of 5 gp41 mAbs ranged from ~1 to 71 nM, whereas binding to vaccine-Env VRC-A and heterologous group M consensus gp140 Envs ranged from ~1 to 23 nM (table S17). None of the gp41-reactive mAbs mediated neutralization using a panel of pseudotyped HIV-1 isolates: neutralization-resistant (tier 2) VRC-A (A.92RW020) and VRC-C (C.97ZA012), neutralization-sensitive (tier 1) B.HxB2 and B.BaL of VRC-B (B.HxB2/BaL-V3), and heterologous neutralization-sensitive (tier 1) B.MN (table S18). Only 1/17 (6%) gp41-reactive mAb captured infectious virions (B.NL4-3) (fig. S5A) or mediated Ab-dependent cell-mediated cytotoxicity (ADCC) (C.1086) (fig. S6). In contrast, one-third (4/12) of vaccine-induced gp120 mAbs neutralized at least 1 neutralization-sensitive HIV-1 isolate (table S19); all neutralizing gp120-reactive Abs targeted the third variable loop (V3) (mAbs DH196, DH449, DH450, and DH452) and neutralized HIV-1 B.BaL and/or B.MN. Of the gp120-reactive mAbs, 2/12 (17%) captured infectious HIV-1 (fig. S5B), and 5/12 (42%) mediated ADCC (fig. S7).

Site of reactivity of vaccine-induced gp41 Abs

We previously reported that gp41-reactive Abs from HIV-1–infected patients are cross-reactive with the 37-kD subunit of bacterial *Escherichia coli* (*E. coli*) RNA polymerase (*2*). We identified a short region of sequence similarity between gp41 and bacterial RNA polymerase with a shared amino sequence of LRAI (amino acid numbering 556–559 in gp41 (PDB: 1AIK) and 43–46 in bacterial RNA polymerase) (PDB: 1BDF) (fig. S8A). A structural alignment of HIV-1 Env gp41 (*26*) and bacterial RNA polymerase (*27*) showed that the α-subunit helices at the RNA polymerase dimer interface were similar to a portion of the gp41 heptad repeat 1 (HR-1) and heptad repeat 2 (HR-2) helices with a 1.30 Å backbone atom root-mean-square deviation (fig. S8, B and C). This analysis raised the hypothesis that the VRC vaccine–induced gp41-reactive Abs may bind to the gp41 postfusion structure.

To determine binding sites for vaccine-induced gp41-reactive mAbs on vaccine VRC-A gp140 protein, we proteolytically cleaved the gp140 protein with trypsin and analyzed the cleavage products on SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gels in Western blot analysis with three VRC-A gp140– and gp41-reactive (DH438, DH440, and DH432) and control mAbs and used liquid chromatography–mass spectrometry (LC-MS). We found a ~25-kD fragment of VRC-A gp140 that was blotted by the gp41, VRC-A gp140, and IM-reactive vaccine-induced mAbs and contained a peptide amino acid sequence (558–567, PDB: 1AIK) AIEAQQHLLQ that placed it in the HR-1 gp41 region overlapping with the LLRAIE gp41 sequence (amino acids 555–560, PDB: 1AIK) of HR-1 that was a putative cross-reactive region with the bacterial IM-protein, RNA polymerase (fig. S9). Thus, one region of VRC-A gp140 bound by HIV-1 gp41-reactive mAbs was in the gp41 HR-1 region.

In order to identify additional candidate proteins in IM that cross-reacted with gp41-reactive Abs, SDS-PAGE gels were run with IM whole-cell lysates (WCLs) and analyzed in Western blots with VRC-A gp140, gp41, and IM-reactive IGHV1-69 mAbs DH438, DH440, and DH432. Analysis of IM sequences in bands reactive with one or more gp41-reactive mAbs revealed 19 candidate IM proteins by LC-MS (fig. S10, a and b). Alignment of candidate IM protein and gp41 sequences demonstrated one bacterial protein, pyruvate-flavodoxin oxidoreductase, that had a sequence similar to that of gp41 HR-1 amino acids 555–560 LLRAIE, with an amino acid sequence in pyruvate-flavodoxin oxidoreductase sequence amino acids 500–505 LLRGIK (fig. S10c). Thus, pyruvate-flavodoxin oxidoreductase is a second candidate protein that may cross-react with HIV-1 gp41-reactive Abs. That this sequence similarity between gp41 and pyruvate-flavodoxin oxidoreductase was the same as found for the tryptic fragment of DH438-bound VRC-A gp140 (fig. S9), as well as the sequence similarity found between gp41 and bacterial RNA polymerase (fig. S8), raised the hypothesis that gp41-reactive Abs induced by the VRC vaccine bind to an HR-1 sequence in the gp41 postfusion conformation. Indeed, vaccine-induced gp41-reactive mAbs DH438, DH432, and DH440 bound to linear peptides containing the LLRAIE HR-1 sequence (table S20), and 16/17 gp41-reactive mAbs bound a gp41 postfusion conformation recombinant protein (*21*) (table S16).

Polyreactivity of HIV-1–reactive mAbs

Binding to non–HIV-1 molecules was used to identify polyreactivity among these HIV-1–reactive Abs (*28*, *29*). Of the gp41-reactive mAbs, 11/17 (65%) reacted with ≥1 of 9 non–HIV-1 host proteins and nucleic acids (table S21 and Fig. 2A); 4/17 (23%) reacted with mitochondrial inner membrane lipid component, cardiolipin (table S21), and 5/17 (29%) reacted with HIV-1 uninfected cells (HEp-2 cells) in an immunofluorescence reactivity assay (table S22 and Fig. 2B). We found DNA prime, rAd5-boost, vaccine-induced gp41-reactive Abs to be reactive with WCLs of anaerobic (14/17, 82%) or aerobic (13/17, 76%) IM. Of gp41-reactive Abs, 47% were reactive with the 37-kD subunit of *E. coli* RNA polymerase that has been shown to cross-react with HIV-1 gp41-reactive Abs (*2*) (Fig. 2C and fig. S11, a and c). In

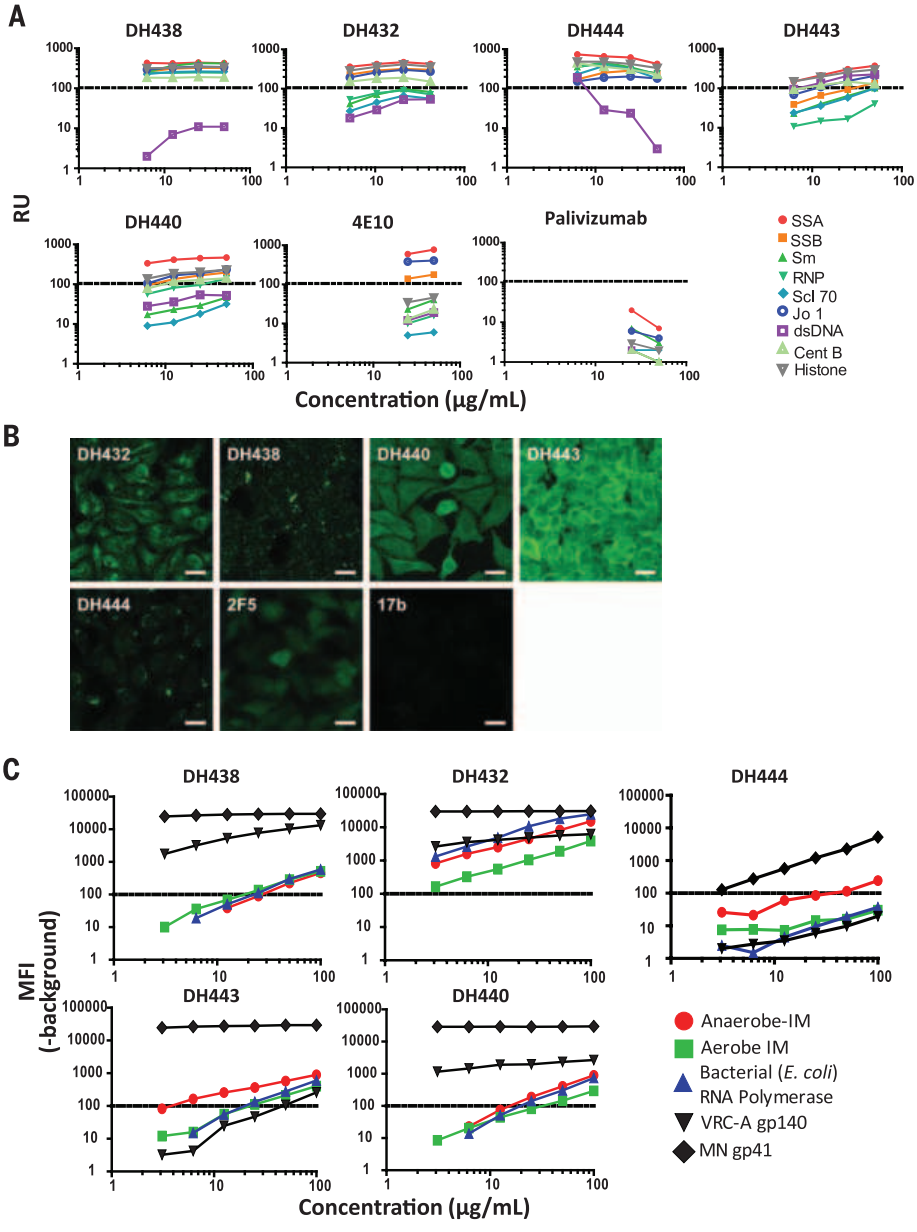


Fig. 2. Polyreactivity of vaccine-induced gp41-reactive mAbs. Five representative polyreactive and gp41-reactive mAbs demonstrated binding with autoantigens [antinuclear Ab (ANA)] (**A**), Hep-2 cells (immunofluorescence staining) (**B**), and anaerobe and aerobe WCLs of IM, bacterial (*E. coli*) RNA polymerase, MN gp41, and vaccine-strain VRC-A gp140 antigens (BAMA) (**C**). (B) All images were taken on an Olympus A×70 fluorescence microscope by using a 40× objective with a SPOT Flex camera. All images were prepared using 50 μg/ml of mAb. Images were acquired for 5 s (2F5 and 17b) or 10 s (DH432, DH438, DH440, DH443, and DH444). Scale bars, 25 μm. (A) and (B) Control Abs: 4E10 (gp41), 2F5 (gp41), 17b (gp41), and palivizumab (RSV). (C) MAb binding was reported as the shift in fluorescence intensity of a population of cells (MFI) versus background versus blank, where background is plate-specific background binding, and blank represents nonspecific sample binding to a negative control bead. Positivity cutoff for binding was 100 MFI, as previously reported (*2*); AbCLL1324 as an IM-reactive control was used to validate IM binding, and gp41 mAb 7B2 that does not react with IM was used as a negative control. Data shown were generated from the same commercially available kits [(A) and (B)], and a single BAMA experiment (C), which was in agreement with Ab-IM reactivity in Western blots (fig. S10 or not shown).

contrast, of the gp120-reactive mAbs, 4/12 (33%) were reactive with ≥1 of 9 host proteins and nucleic acids (table S21), 1/12 (8%) with cardiolipin (table S21), 1/12 (8%) with HEp-2 cells (table S22), 8/12 (67%) with anaerobic IM-WCLs, 5/12 (42%) with aerobe IM-WCLs, and 2/12 (17%) with *E. coli* RNA polymerase (fig. S11, b and c). Collectively, DNA prime–rAd5 boost vaccine-induced gp41-reactive mAbs were more polyreactive than gp120-reactive mAbs ($P < 0.003$, Cochran-Mantel-Haenszel test) (Table 2).

NGS of prevaccine IgHV repertoires

The isolation of Abs from HIV-1–infected and naïve individuals that cross-reacted with gp41 and IM-WCLs led to the hypothesis that a pool of B cells exist that are cross-reactive with Env gp41 and IM (*2*, *3*). To determine whether these types of naïve B cells can respond in the setting of HIV-1 Env gp140 vaccination, we performed immunoglobulin variable heavy chain NGS on prevaccination PBMC–derived RNA samples from the eight vaccinees from which postvaccine Env-specific memory B cells were isolated. The immunoglobulin repertoire of *IGHV1-IGHV6* gene families for IgA, G, and M isotypes was probed in prevaccination blood B cells (table S24). We used the V_HDJ_H DNA sequences of all postvaccination Abs with unique V_HDJ_H rearrangements from each of the eight vaccinees to search their prevaccination B cell repertoire for V_HDJ_H rearrangements that belonged to members of the same clonal lineages. We tested the relatedness among sequences by fitting alternative phylogenetic models in which the sequences were or were not hypothesized to share a common ancestor. In both models, we computed a Bayesian average over ancestral sequences; two such ancestors in the model for similar sequences derived from different B cells, one ancestor for the model for clonally related sequences. In the latter case, we also averaged over all unobserved intermediates as well. The model with the larger summed likelihood was selected.

From vaccinee 082-003, we found a prevaccination IgM V_HDJ_H rearrangement (DH477) that was clonally related to a postvaccination gp41-reactive IgG1 Ab, DH476. Direct comparison of these sequences revealed shared junctional sequences, identical HCDR3 length, and 85% HCDR3 nucleotide sequence homology (Fig. 3, A and B, and fig. S12), which demonstrated that prevaccination DH477 and postvaccination DH476 Abs came from the same B cell precursor. In addition to being reactive with all three recombinant vaccine-Envs, the postvaccination Ab DH476 was reactive with rgp41 and with both anaerobe and aerobe IM-WCLs (Fig. 3, C, D, and F). The prevaccine clonal lineage member DH477 was an unmutated IgM Ab complemented by the postvaccine DH476 light chain. Prevaccine Ab DH477 reacted with vaccine strain Env, MN gp41, and host and/or environmental antigens, including IM (Fig. 3, C to F, and tables S25 and S26). The prevaccine Ab DH477 had greater reactivity to IM than the postvaccine Ab DH476 (Fig. 3F) and only bound VRC-A Env surface-expressed on human embryonic kidney 293 (HEK293) cells (Fig. 3E) but did not bind VRC-A, -B, or -C gp140 recombinant proteins (Fig. 3C). Reactivity with IM-WCLs decreased from prevaccination Ab DH477 to postvaccination Ab DH476, whereas increased binding of vaccine Env gp140

Fig. 3. Characteristics of clonally related Abs. Pre- (DH477) and post- (DH476) vaccine clonally related Abs found in vaccinee 082-003 had the same V_HDJ_H recombination. IGHV, IGHD, and IGHJ Ab segments were statistically inferred (**A** and **B**); nucleotides in the 3′ end of the V-gene and 5′ end of the J-gene shown with rearrangement junctions (gray-shaded nucleotides—NN1, NN2), and HCDR3 (underlined nucleotides) indicated. The lower-case letters show nucleotides that have been removed from the germline gene during V_HDJ_H rearrangement. Pre- (DH477) and post- (DH476) vaccine heavy chain sequences were paired with the natural light chain (IGκV3-20*01, IGκJ1*01, 1.7% mutated nt, 10 amino acid CDR3) of the postvaccine Ab DH476. (**C**) Ab binding to recombinant HIV-1 Env gp140, 5-Helix gp41, and MN gp41 proteins was determined by surface plasmon resonance (SPR) analysis, and (**D**) the Ab binding K_d was determined by rate constants or steady-state analysis measurements. (**E**) Binding of DH477 to VRC-A gp140 expressed on the surface of 293i cells. Anti-V3 mAb (19B) and a representative gp41-reactive Ab (DH440) from this study, both of which bind recombinant VRC-A gp140, were used as positive controls, whereas palivizumab (RSV Ab) was used as a negative IgG1 control. (**F**) The cross-reactivity of pre- (DH477) and post- (DH476) vaccine Abs with intestinal microbiota (IM) whole cell lysates (WCLs) was determined by means of Western blot and SPR as previously described (*2*). DH438 (gp41-reactive Ab) and palivizumab were used as positive and negative control Abs, respectively, for Western blots, and the arrows indicate candidate IM antigens bound by DH476 and/or DH477. In SPR analysis, IM WCLs were diluted in PBS buffer as indicated and injected over mAbs captured on an IgFc-specific Ab immobilized sensor surface. Specific binding of the lysate proteins to the mAbs in SPR were measured after subtraction of nonspecific binding to HA-specific mAb CH65. Abbreviations: nt, nucleotides; aa, amino acid. Data shown in (C) to (F) were representative of duplicate experiments, except SPR screen of DH477 and DH476 for binding rgp140 proteins that was performed once (C).

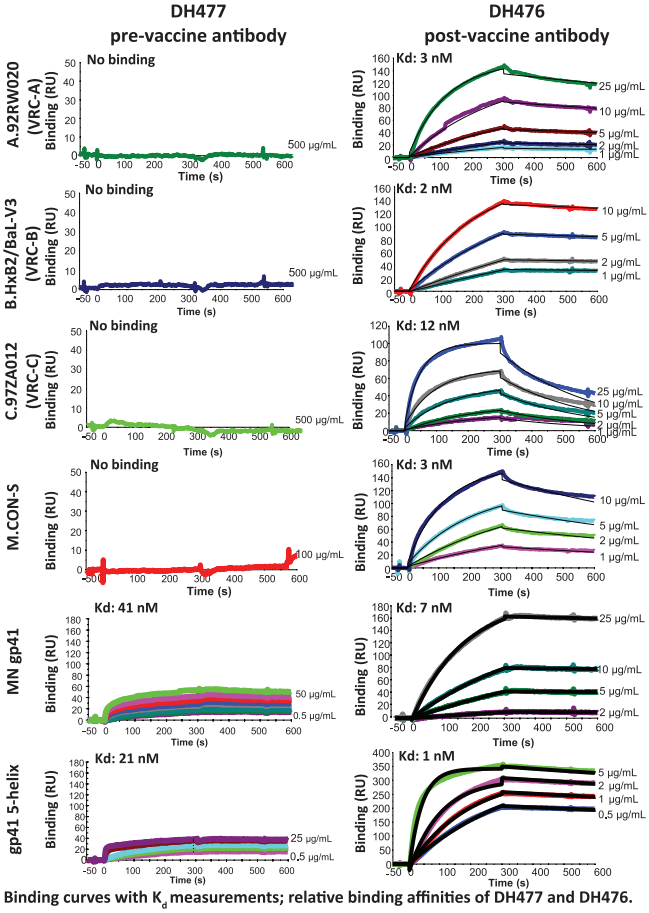
A Immunogenetics

PTID	HVTN Trial	Vaccination timepoint	mAb ID	VH	DH	JH	VH mutation frequency (%), nt	HCDR3 length, aa	Ig Isotype
082-003	082	PRE	DH477	1-69*06	3-9*01	4-1*02	0	11	M
		POST	DH476	1-69*06	3-9*01	4-1*02	5.3	11	G

B DH477/DH476_VHDJH junctions

Antibodies/ V_HDJ_H	$V_H(3')$	NN1	D_H	NN2	$J_H(5')$
DH477 (PRE)	TGTGCGAGAG	CGG	ACGATATTT	GGCCTC	TTGACTACTGG
DH476 (POST)	-------C--	G--	-T---T-G-	------	----------
IGHV1-69*06	----------	a			
IGHD3-9*01	........gt	att	---------	tgactg	gttattataac
IGHJ4-1*02	....	...	....	actact	---------- (286-324)

C SPR analysis of DH477 and DH476 binding profile



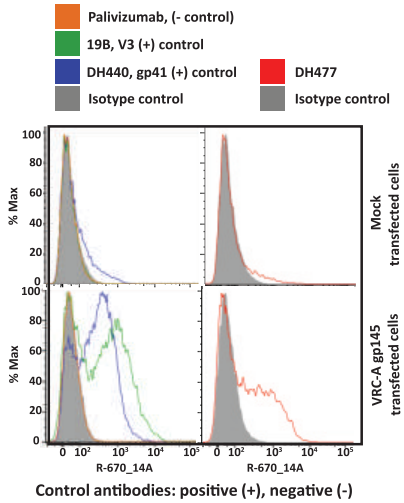
Binding curves with K_d measurements; relative binding affinities of DH477 and DH476.

D DH477/DH476 Antibody binding kinetics

HIV-1 Envelopes	DH477 (pre-vaccine) k_a, $M^{-1}s^{-1}$	k_d, s^{-1}	K_d, nM	DH476 (post-vaccine) k_a, $M^{-1}s^{-1}$	k_d, s^{-1}	K_d, nM
A.92RW020	*nb*	*nb*	*nb*	$13x10^4$	$4x10^{-4}$	3
B.HxB2/BaL-V3	*nb*	*nb*	*nb*	$20x10^4$	$3x10^{-4}$	2
C.97ZA012	*nb*	*nb*	*nb*	$28x10^4$	$34x10^{-4}$	12
M.CON-S	*nb*	*nb*	*nb*	$32x10^4$	$10x10^{-4}$	3
MN gp41	---	---	41*	$10x10^3$	$7x10^{-5}$	7
gp41 5-helix	---	---	21*	$2x10^5$	$2x10^{-4}$	1

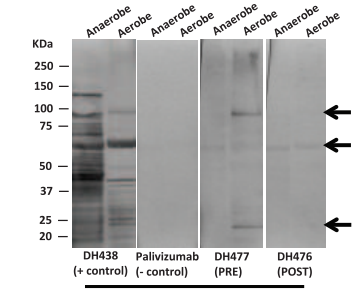
*K_d determined by steady state analysis. nb, no binding.

E DH477 binding of VRC-A gp145 Env

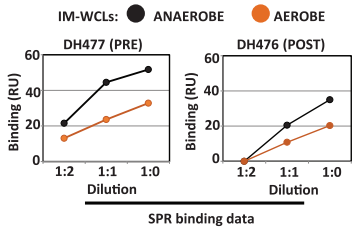


Control antibodies: positive (+), negative (-)

F DH477/DH476 Cross-reactivity profile



SDS-PAGE reduced gels (western blot)



SPR binding data

Table 2. Polyreactivity of DNA prime–rAd5 boost vaccine–induced gp41-reactive and gp120-reactive mAbs. Abs were screened for cross-reactivity with anaerobe and aerobe WCLs of IM from human stool and recombinant bacterial (*E. coli*) RNA polymerase (BAMA); host antigens (AtheNA); HEp-2 cells (IFA, indirect fluorescence Ab assay, Zeuss Scientific); and cardiolipin (QUANTA Lite ACA IgG III). *P* values were generated by means of Fisher's exact test (SAS v9.3) or, for all Abs, Cochran-Mantel-Haenszel test controlling for antigen (SAS v9.3).

Antigens	gp41-reactive Abs	gp120-reactive Abs	*P* value
Anaerobe IM	14/17 (82%)	8/12 (67%)	0.40
Aerobe IM	13/17 (76%)	5/12 (42%)	0.12
Bacterial (*E. coli*) RNA polymerase	9/17 (53%)	2/12 (17%)	0.06
AtheNA autoantigen panel	11/17 (65%)	4/12 (33%)	0.14
HEp-2 cells (IFA)	5/17 (29%)	1/12 (8%)	0.35
Cardiolipin	4/17 (24%)	1/12 (8%)	0.63
All antigens tested, gp41-reactive versus gp120-reactive Abs			0.003

and postfusion forms of gp41 was associated with vaccine-Env–induced affinity maturation (Fig. 3, C to F). Thus, vaccine-Env gp41-reactive Ab response can indeed arise from a pool of IM-gp41 cross-reactive B cells that undergo postvaccination affinity maturation.

Discussion

In this study, we found that the DNA prime–rAd5 boost HIV-1 vaccine induced a non-neutralizing dominant Env gp41-reactive Ab response derived from a polyreactive B cell pool. These findings have important implications for HIV-1 vaccine design. The isolation of gp41 and IM-reactive B cells from uninfected individuals (*2*, *3*) raised the hypothesis that the HIV-1 gp41-reactive Ab response in HIV-1 infection may in part result from HIV-1 gp41 stimulating a preinfection pool of B cells cross-reactive with IM and Env. We confirmed this hypothesis by the identification of a prevaccination IM, VRC-A gp140, and gp41-reactive Ab that was clonally related to a vaccine-induced Env-reactive Ab (Fig. 3).

The prevaccination gp41-reactive Ab DH477 was an unmutated IgM Ab, which suggested that it represented the receptor of a naïve B cell. Although the great majority of human IgM^+, IgD^+ B cells are naïve B cells that are unmutated, a minority of IgM^+, IgD^- memory B cells are unmutated as well (*30*, *31*). Thus, DH477 could have been derived from an IgM^+, IgD^+ naïve B cell, but we cannot rule out as well that DH477 was derived from an unmutated IgM^+, IgD^- memory B cell. That the pre- (DH477) and post- (DH476) vaccine Abs reacted with IM-WCLs, gp41, and vaccine-Env proteins, suggested that IM or other environmental antigens may stimulate gp41-reactive B cells before vaccination that can react with the vaccine Env but does not prove that all such Abs arose in this manner. B cell development in mice has been demonstrated to occur in the intestinal mucosa and is regulated by extracellular signals from IM that influence the gut immunoglobulin repertoire (*32*). Our findings also implicate IM imprinting of B cell development in humans that can affect the quality of an HIV vaccine Ab response.

Here, we demonstrated that an HIV-1 Env-gp140 vaccine containing both gp120 and gp41 primarily induced a gp41-dominant Ab response, whereas previous gp120-only vaccines in humans (*14*) and rhesus macaques (*15*) induced a robust gp120-reactive memory B cell response. It has been demonstrated that B cells with higher affinity for an antigen can outcompete other B cells with lower affinity for the same antigen (*33*–*35*), and B cells in vitro with similar, but slightly different, B cell receptors can compete for antigen binding (*36*). Thus, our data raise the additional hypothesis that the dominant non-neutralizing IM cross-reactive gp41-reactive Abs outcompeted gp120-reactive Abs.

Of interest, the gp41-reactive Ab response was predominantly restricted to *IGHV1-69* gene segment usage. Previously, IGHV1-69 has been noted for influenza stalk (*20*, *37*), HIV-1 gp41 hydrophobic pocket (*21*), and membrane-proximal region Ab responses (*38*). That the gp41 Ab response constituted predominantly IGHV1-69 memory B cells that used the 54L-HCDR2 IGHV1-69 alleles suggested that the 54L-HCDR2 *IGHV1-69* gene allele variant was the predominant IGHV better able to structurally recognize the dominant gp41 epitope expressed by the DNA prime–rAd5 boost vaccine. However, it is important to note that prevaccine Ab DH477 used the 54F-HCDR2 variant of IGHV1-69 (Fig. 3 and fig. S12), whereas the clonally related postvaccine Ab DH476 used the 54L IGHV1-69 variant (table S5 and fig. S12). Thus, a second mechanism of 54L VH1-69 usage was via somatic hypermutations to recognize gp41 epitopes presented by gp140 immunogens in the DNA prime–rAd5 boost vaccine.

One hypothesis for vaccine optimization is to use a native trimer containing gp41 to attempt to induce broadly reactive neutralizing Abs (bnAbs) (*39*, *40*). These immunogens contain the recently described gp41-gp120 conformational bnAbs epitopes (*41*–*44*) but not the membrane-proximal portion of gp41-neutralizing epitopes (*45*, *46*). The Env insert in the VRC Env DNA-Ad5 vaccine had a partial deletion in the C-C loop of gp41, unlike the native Env. Although it has been reported that another HIV-1 Env vaccine with the gp41 C-C loop intact has also induced high levels of gp41-reactive Abs (*47*), this vaccine has not been studied in an efficacy trial, nor have its vaccine-induced gp41-reactive mAbs been characterized, not even for environmental or IM cross-reactivity. In nonhuman primate studies, gp41-reactive plasma Abs have been reported to be both protective (*48*, *49*) and nonprotective (*50*, *51*). Defining the B cell compartment from which polyreactive gp41-reactive B cells originate and determining how they are regulated are important areas of future study. The low proportion of gp120 Abs does not exclude the possibility that additional Abs were generated against other forms of the ectodomain, for example, conformational epitopes. Future studies will evaluate this possibility.

Thus, as in HIV-1 infection, Env vaccination can induce polyreactive, non-neutralizing gp41-reactive Ab repertoire responses from preexisting B cells that can be cross-reactive with IM. Because the B cell repertoire can be imprinted at birth by IM antigens (*32*), our data raise the hypothesis that neonatal immunization with HIV-1 Env may be able to imprint the B cell repertoire to respond to Env antigenic sites that may otherwise be subdominant or disfavored, including Env broadly neutralizing Ab epitopes (*52*, *53*). In this regard, Goo et *al.* recently reported that HIV-1–infected infants can make broadly neutralizing Abs, and in some cases, such Abs arise within the first year of life (*54*).

Finally, if similar diversion toward dominant polyreactive gp41-reactive Abs is found with other gp41-containing vaccine regimens with intact gp41, the data in this study suggest that one region of gp41 in vaccine immunogens that might be considered for deletion or modification to avoid IM-gp41 cross-reactivity in the setting of vaccination includes amino acids in the gp41 HR-1 region.

Materials and methods

Clinical trial samples

Vaccine-induced Ab repertoires were studied 4 weeks after final vaccination (rAd5 boost) in plasma and blood-derived memory B cells of four HVTN 204 (*5*), and four HVTN 082 (Protocol HVTN 082; DIADS Document ID 10771) trial participants, as well as in plasma only from an additional 40 HVTN 505 vaccine recipients (*11*). Vaccinees were HIV-1 seronegative adults enrolled in the phase Ib HVTN 082 (Protocol HVTN 082; DIADS Document ID 10771) and phase II HVTN 204 (Protocol HVTN 204; BB IND 12326 HELD BY DAIDS) (*5*) trials that studied the DNA prime–rAd5 boost vaccine. From $n = 480$ (HVTN 204), the four participants chosen for our study displayed plasma binding reactivity with vaccine strain Envs, and/or neutralization of clade C MW965 HIV-1 isolate. From $n = 8$ (HVTN 082, four twin pairs), we selected one individual from each twin pair to study genetically different subjects. All study samples were obtained by informed consent, and all studies were approved by the Duke University Institutional Review Board.

Flow cytometry memory B cell single-cell sorting

Single-cell isolation of memory B cells decorated with both AlexaFluor 647 and Brilliant Violet 421–tagged HIV-1 VRC-A (A.92RW020), VRC-B (B.HxB2/BaL-V3), VRC-C (C.97ZA012), or CON-S Env gp140s or gp120s was performed using a fluorescence-activated cell sorter, either FACSAria or FACSAria II (BD Biosciences, San Jose, CA), and the flow cytometry data were analyzed using FlowJo (Treestar, Ashland, OR) (*13*, *14*, *55*).

Ab binding and epitope mapping

Recombinant mAbs were screened for binding specificities to multiclade HIV-1 gp120 Envs and clade B.MN gp41 (Product 10911, ImmunoDX, Woburn, MA) by means of standard ELISA (*56*). Gp120-reactive Abs bound both Env gp140 and gp120 recombinant proteins, but gp41-reactive Abs bound recombinant MN gp41 protein or differentially to CON-S and/or VRC A Envs (CON-S gp140+, CON-S gp120−; VRC-A gp140+, VRC-A gp120−). MAbs were epitope-mapped on multiclade consensus and primary isolates gp140 overlapping peptide sets, including HIV-1 MN and B.CON gp140 obtained from the NIH AIDS Reagent Repository, by means of ELISA and linear peptide array (*57*). The binding Ab multiplex assay (BAMA)—a standardized custom binding Ab multiplex assay—was used to determine reactivity of serum and recombinant mAbs with antigens (*1*). Prevaccinated serum samples and a respiratory syncytial virus (RSV)–specific mAb (palivizumab) were used to establish plasma and mAb binding cutoffs, respectively, in BAMA. For SPR analysis, Env gp140 (A.92RW020, B.HxB2/Bal, C.97ZA012, or M.CON-S), 5-helix gp41 (*21*), and MN gp41 proteins (Product 10911, ImmunoDX) binding was measured by injecting Env proteins at varying concentration (0.5 to 500 μg/ml) over each mAb captured on human IgFc–specific immobilized Ab (Millipore) on a CM5 sensor surface as previously described (*58*). Nonspecific binding to the control palivizumab mAb and signal drifts from phosphate-buffered saline (PBS) were used for double referencing and to measure specific binding responses. Rate constants for association and dissociation (k_a and k_d, respectively) were measured by global curve fitting to a 1:1 Langmuir model as described previously (*58*). For gp41 proteins (MN gp41 and 5-helix) binding to prevaccine Abs, rate constants could not be reliably measured because of saturation of binding responses at relatively lower antigen concentrations, and values for the dissociation constant (K_d) were measured using steady-state analysis. Postvaccine Abs did not show this limitation when binding to either MN gp41 or 5-helix gp41 proteins (Fig. 3).

Ab staining of cell surface–expressed HIV-1 VRC-A gp145

For the surface expression of VRC-A gp145, HEK293i cells were transfected with pHV130770 plasmids using ExpiFectamine 293 Transfection Kits (Life Technologies, A14525, Grand Island, NY) following the manufacturers' protocol. The cells were washed with four volumes of room-temperature phosphate-buffered saline (DPBS) (Life Technologies 10010-023) and centrifuged (500*g* for 5 min). Cells (10^6) were resuspended in 80 μl of staining buffer (1% BSA in DPBS) to which were added 20 μl of diluted Abs (1 μg per reaction). After incubation (1 hour at 4°C) with shaking, cells were washed and centrifuged (500*g*). Cells were incubated with either 20 μl isotype-APC control (BD cat. 555751) or 20 μl human IgG-APC–specific (BD cat 550931) for 30 to 45 min at 4°C, protected from light. After washing and fixation, cells were analyzed by flow cytometer (BD LSRFortessa).

Ab polyreactivity

Vaccine-induced polyreactivity of Abs was assessed by using commercially available kits (AtheNA Multi-Lyte System and HEp-2 cell immunofluorescence assay, Zeus Scientific; anticardiolipin ELISA - QUANTA Lite ACA IgG III, Inova Diagnostics, San Diego, CA) (*3*). Reactivity with WCLs of anaerobic and aerobic IM extracts of stool specimens (*3*) was determined by using BAMA and Western blot (WB), as described previously (*2*). For WB analysis, 20 μg/ml of mAbs was used to test reactivity with 100 μg of IM WCLs. SPR analysis of IM WCL samples was performed on a BIAcore 3000 instrument (GE Healthcare), as previously described (*2*), with some modifications. IM WCL samples were spun down to remove aggregates, and each lysate sample (total protein concentrations of 7.87 and 7.97 mg/ml for aerobic and anaerobic lysates, respectively) was diluted in PBS buffer (1:1 or 1:2) and injected over each mAb captured over a human IgFc–specific Ab immobilized on a C1 sensor chip. Specific binding of the lysate proteins to the mAbs was measured after subtraction of nonspecific binding to HA-specific mAb CH65, and binding responses were measured at post-injection report point (10 s after injection).

Proteomic analysis of IM proteins and HIV-1 VRC-A gp140

SDS-PAGE gels were run with IM proteins, and Western blot analysis was performed with three VRC-A gp140 and MN gp41-reactive, IM cross-reactive IGHV1-69–using mAbs: DH440, DH432, and DH438. Bands were cut from gels that were reactive with one or more of these mAbs in Western blot, and we performed proteomic analysis by LC-MS of tryptic fragments of these protein bands to identify the reactive protein. One of these bands (An1) was recognized by all three gp41-reactive mAbs, and an identified protein in the WB-reactive band An1 (fig. S10) was bacterial pyruvate-flavodoxin oxidoreductase (National Center for Biotechnology Information, NIH, accession no: WP_022497386).

In order to directly determine where on the VRC-A gp140 protein the vaccine-induced gp41 mAbs bound, we proteolytically cleaved the gp140 protein with trypsin and then analyzed the cleavage products by WB analysis. One of the reactive bands (~25 kD) that blotted with the gp41-reactive mAb tested was identified by LC-MS proteomic analysis after reduction and alkylation.

Sequence and structural similarities of HIV-1 Env gp41 and bacterial proteins

Alignment of bacterial protein identified by LC-MS with gp41 Env sequences was assessed by BLAST analysis of the VRC-A, VRC-B, and VRC-C gp140 and MN gp41 amino acid sequences against sequences of 19 candidate proteins identified in IM by LC-MS. We used a cutoff E-value of <1 to allow for short stretches of sequence matches. The BLAST search returned only one similar region in pyruvate flavodoxin oxidoreductase, where sequence similarity was observed below the E-value cutoff between bacterial proteins and the gp41 ectodomain. The H1 helices at the homodimer interface of the α subunit of RNA polymerase (PDB: 1BDF) were structurally aligned to the six-helix bundle conformation of gp41 (PDB: 1AIK), by using an implementation of the Kabsch algorithm.

Ab functional characterization

Purified mAbs were screened for neutralization by means of TZM-bl assay (*59*, *60*), ADCC (*61*, *62*), and infectious virion capture (*63*) of HIV-1 isolates.

PCR isolation of heavy and light chain genes

Heavy (*IGHV*) and light (*IGKV*, *IGLV*) chain genes were isolated by means of single-cell PCR (*13*, *64*), and the sequences were computationally determined as described (*65*–*67*). Inferences of the genetic features of the immunoglobulin *IGHV*, *IGKV*, and *IGLV* sequences were previously described (*68*, *69*).

Expression of IGHV, IGKV, and IGLV chains as full-length IgG1 rmAbs

Plasmids encoding the *IGHV*, *IGKV*, and *IGLV* genes were generated and used for recombinant monoclonal Ab (rmAb) production in human embryonic kidney cell lines (ATCC, Manassas, VA) (*13*, *38*), by means of small-scale transfection and as purified mAbs in larger quantities (*70*, *71*). Purified rmAbs were dialyzed against PBS, analyzed, and stored at 4°C.

Clonal lineage determination

Sequences were subject to statistical analysis for lineage membership. Briefly, sequences were organized into clans, by definition sharing inferred *IGHV* and *IGHJ* genes and CDRH3 length. Within these clans, we used agglomerative clustering using a Bayesian phylogenetic merit function as follows. To calculate the score for the hypothesis that a given sequence is a member of a lineage L, we compute the posterior likelihoods under the phylogenetic hypotheses that they are and are not members of the same clone. The Bayes factor for this comparison is the objective function for clustering. Sequences were taken in succession and placed into the lineage (including an empty lineage, i.e., founding a new lineage) that maximizes the objective function. The automated inference was followed up by visual inspection of the DNA sequence alignments for confirmation.

Statistical analysis

Statistically significant differences in the profiles of gp120-reactive and gp41-reactive Abs were determined by means of various tests (as indicated for each table and figure) in SAS software (SAS v9.3). A $P < 0.05$ determined statistical significance.

NGS

PBMC-extracted RNA was used to generate cDNA amplicons for pyrosequencing (Illumina). RNA isolated from vaccinee PBMCs was separated into two equal aliquots before cDNA production; cDNA amplification and NGS were performed on both aliquots as independent samples. *IGHV* genes were amplified by means of a modification of the processes previously described (*13*). The reverse transcription (RT) reaction was carried out in 30-μl reaction mixtures at 55°C for 1 hour after addition of 50 units/reaction of Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA), 40 units/reaction of RNaseOUT (Invitrogen), 25 mM deoxynucleotide triphosphates (dNTPs) (Invitrogen), 5× first strand buffer (Invitrogen), 0.1 mM DTT (Invitrogen), and 25 μM human IgA, IgG, and IgM constant region primers as previously described (*13*). After cDNA synthesis, *IGHV1-IGHV6* genes were amplified separately for IgA, IgG, and IgM isotypes by two rounds of PCR in 96-well PCR plates in 50 μl reaction mixtures. The first-round of PCR contained 5 μl of RT reaction products, 1 unit of iProof DNA polymerase (Biorad), 10 μl 5× iProof GC buffer (Biorad), 10 mM dNTPs (Invitrogen), 50 mM $MgCl_2$ (Biorad), and 25 μM of IgA, IgG, or IgM constant region primers and sets of IGHV-IGHV6 variable region primers (table S23). The first round of PCR was performed at 98°C ×1 min followed by 25 cycles of 98°C ×15 s, 60°C ×15 s, 72°C ×35 s, and one cycle at 72°C ×7 min. First-round PCR products were purified by using a QIAquick PCR purification kit (Qiagen) and eluted into 50 μl of DNase/RNase-free distilled water. Nested second-round PCR was performed in 50 μl of reaction mixture with 5 μl of purified first-round PCR product, 1.25 units/reaction Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), 10 mM dNTPs, and 5 μl of Nextera index kit barcode-tagged primers (Illumina). During the second round of nested PCR, the IGHV1-IGHV6 primers were amplified in separate reaction mixes for each variable region primer. The second round of PCR was performed at 94°C ×2 min followed by three cycles of 94°C ×15 s, 55°C ×30 s, 68°C ×30 s; seven cycles of 94°C ×15 s, 60°C ×30 s, 68°C ×30 s; and one cycle at 68°C ×10 min. Samples of IGHV chain PCR products were analyzed on 2% agarose gels before total sample purification by means of gel extraction (QIAquick gel extraction kit; Qiagen) and eluted into 25 μl of DNase/RNase-free distilled water. IGHV1-IGHV6 cDNA amplicons (5 μl) were pooled for each immunoblobulin isotype. Pooled IgA, IgG, and IgM cDNA amplicons were quantified using the KAPA SYBR FAST qPCR kit (KAPA Biosystems). IgA, IgG, and IgM samples at 4 nM were selected for NGS. IgA, IgG, and IgM samples were denatured with NaOH (1×) and then mixed with hybridization buffer. Denatured samples were mixed with 35% (by volume) PhiX control/nonrelevant DNA (Illumina) in a final reaction mixture, from which 600 μl were loaded to Illumina MiSeq kit v3 for 600 cycles of PCR amplification. We have previously shown that the primer set used here does not induce primer IGHV bias (*3*).

NGS-generated sequences were processed and analyzed computationally according to the following protocol; Illumina reads were discarded if they did not meet the following criteria: ≥30 Phred quality score (corresponding to 99.9% base call accuracy) for at least 95% of all base positions in the sequence. Primer sequences were removed from the reads, and only unique sequences were retained. The unique sequences were then analyzed using the Cloanalyst software suite (*72*) to infer the IGHV rearrangements (V_HDJ_H) and to determine clonal relatedness. Illumina NGS was performed independently on duplicate samples (RNA samples split into two aliquots for independent cDNA production, amplification, and deep sequencing). Illumina runs were performed with independent Illumina kits. For the purpose of gene counting, only sequences that appeared in both duplicate runs (no mismatches in the coding region) were used (tables S12 and S24). For the identification of prevaccine lineage members, replication was not required. The statistical methods we used are not sensitive to random sequencing error, and the likelihood that sequencing errors would produce artifactual apparent lineage membership in an otherwise unrelated sequence is negligible.

We used NGS to study the prevaccination samples of all eight vaccinees in the study, and we found three vaccinees whose prevaccination samples had Abs that were in the same Ab clan (same V_HDJ_H and HCDR3 length and similarity), but two of these prevaccination Abs had ambiguities in their junctional sequences and shared mutations versus allelic differences, such that a definitive conclusion could not be reached that the prevaccination and postvaccination V_HDJ_H originated from the same B cell. However, in the case of the prevaccination DH477 and postvaccination DH476 Abs, from the shared mutations and junctional sequences, we could definitively state that they arose from the same B cell. In order to investigate the binding specificities of the pre- (DH477) and post (DH476) vaccine Abs, pre- and postvaccine V_H chains were paired with the natural light chain of the postvaccine Ab.

REFERENCES AND NOTES

1. G. D. Tomaras *et al*., Initial B-cell responses to transmitted human immunodeficiency virus type 1: Virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82**, 12449–12463 (2008). doi: 10.1128/JVI.01708-08; pmid: 18842730
2. A. M. Trama *et al*., HIV-1 envelope gp41 antibodies can originate from terminal ileum B cells that share cross-reactivity with commensal bacteria. *Cell Host Microbe* **16**, 215–226 (2014). doi: 10.1016/j.chom.2014.07.003; pmid: 25121750
3. H. X. Liao *et al*., Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *J. Exp. Med.* **208**, 2237–2249 (2011). doi: 10.1084/jem.20110363; pmid: 21987658
4. A. T. Catanzaro *et al*., Phase I clinical evaluation of a six-plasmid multiclade HIV-1 DNA candidate vaccine. *Vaccine* **25**, 4085–4092 (2007). doi: 10.1016/j.vaccine.2007.02.050; pmid: 17391815
5. G. J. Churchyard *et al*., A phase IIA randomized clinical trial of a multiclade HIV-1 DNA prime followed by a multiclade rAd5 HIV-1 vaccine boost in healthy adults (HVTN204). *PLOS ONE* **6**, e21225 (2011). doi: 10.1371/journal.pone.0021225; pmid: 21857901
6. H. Kibuuka *et al*., A phase 1/2 study of a multiclade HIV-1 DNA plasmid prime and recombinant adenovirus serotype 5 boost vaccine in HIV-Uninfected East Africans (RV 172). *J. Infect. Dis.* **201**, 600–607 (2010). doi: 10.1086/650299; pmid: 20078213
7. W. Jaoko *et al*., Safety and immunogenicity study of Multiclade HIV-1 adenoviral vector vaccine alone or as boost following a multiclade HIV-1 DNA vaccine in Africa. *PLOS ONE* **5**, e12873 (2010). doi: 10.1371/journal.pone.0012873; pmid: 20877623
8. B. S. Graham *et al*., Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 DNA candidate vaccine. *J. Infect. Dis.* **194**, 1650–1660 (2006). doi: 10.1086/509259; pmid: 17109336
9. A. T. Catanzaro *et al*., Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 candidate vaccine delivered by a replication-defective recombinant adenovirus vector. *J. Infect. Dis.* **194**, 1638–1649 (2006). doi: 10.1086/509258; pmid: 17109335
10. R. A. Koup *et al*., Priming immunization with DNA augments immunogenicity of recombinant adenoviral vectors for both HIV-1 specific antibody and T-cell responses. *PLOS ONE* **5**, e9015 (2010). doi: 10.1371/journal.pone.0009015; pmid: 20126394
11. S. M. Hammer *et al*., Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *N. Engl. J. Med.* **369**, 2083–2092 (2013). doi: 10.1056/NEJMoa1310566; pmid: 24099601
12. H. X. Liao *et al*., A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. *Virology* **353**, 268–282 (2006). doi: 10.1016/j.virol.2006.04.043; pmid: 17039602
13. H. X. Liao *et al*., High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. *J. Virol. Methods* **158**, 171–179 (2009). doi: 10.1016/j.jviromet.2009.02.014; pmid: 19428587
14. M. A. Moody *et al*., HIV-1 gp120 vaccine induces affinity maturation in both new and persistent antibody clonal lineages. *J. Virol.* **86**, 7496–7507 (2012). doi: 10.1128/JVI.00426-12; pmid: 22553329
15. K. Wiehe *et al*., Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity* **41**, 909–918 (2014). doi: 10.1016/j.immuni.2014.11.014; pmid: 25526306
16. S. D. Boyd *et al*., Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol.* **184**, 6986–6992 (2010). doi: 10.4049/jimmunol.1000445; pmid: 20495067
17. E. H. Sasso, T. Johnson, T. J. Kipps, Expression of the immunoglobulin VH gene 51p1 is proportional to its germline gene copy number. *J. Clin. Invest.* **97**, 2074–2080 (1996). doi: 10.1172/JCI118644; pmid: 8621797
18. 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010). doi: 10.1038/nature09534; pmid: 20981092
19. R. A. Lerner, Rare antibodies from combinatorial libraries suggests an S.O.S. component of the human immunological repertoire. *Mol. Biosyst.* **7**, 1004–1012 (2011). doi: 10.1039/c0mb00310g; pmid: 21298133
20. R. Xu *et al*., A recurring motif for antibody recognition of the receptor-binding site of influenza hemagglutinin. *Nat. Struct. Mol. Biol.* **20**, 363–370 (2013). doi: 10.1038/nsmb.2500; pmid: 23396351
21. M. A. Luftig *et al*., Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nat. Struct. Mol. Biol.* **13**, 740–747 (2006). doi: 10.1038/nsmb1127; pmid: 16862157
22. C. Sabin *et al*., Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41. *PLOS Pathog.* **6**, e1001195 (2010). doi: 10.1371/journal.ppat.1001195; pmid: 21124990
23. K. K. Hwang *et al*., IGHV1-69 B cell chronic lymphocytic leukemia antibodies cross-react with HIV-1 and hepatitis C virus antigens as well as intestinal commensal bacteria. *PLOS ONE* **9**, e90725 (2014). doi: 10.1371/journal.pone.0090725; pmid: 24614505
24. M. A. Moody *et al*., H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin

antibodies than influenza vaccination. *PLOS ONE* **6**, e25797 (2011). doi: 10.1371/journal.pone.0025797; pmid: 22039424
25. N. Darzentas *et al*., A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: Molecular and computational evidence. *Leukemia* **24**, 125–132 (2010). doi: 10.1038/leu.2009.186; pmid: 19759557
26. D. C. Chan, D. Fass, J. M. Berger, P. S. Kim, Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997). doi: 10.1016/S0092-8674(00)80205-6; pmid: 9108481
27. G. Zhang, S. A. Darst, Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science* **281**, 262–266 (1998). doi: 10.1126/science.281.5374.262; pmid: 9657722
28. H. Mouquet, M. C. Nussenzweig, Polyreactive antibodies in adaptive immune responses to viruses. *Cell. Mol. Life Sci.* **69**, 1435–1445 (2012). doi: 10.1007/s00018-011-0872-6; pmid: 22045557
29. B. F. Haynes *et al*., Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308**, 1906–1908 (2005). doi: 10.1126/science.1111781; pmid: 15860590
30. U. Klein, R. Küppers, K. Rajewsky, Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood* **89**, 1288–1298 (1997).pmid: 9028952
31. U. Klein, R. Küppers, K. Rajewsky, Human IgM+IgD+ B cells, the major B cell subset in the peripheral blood, express V_K genes with no or little somatic mutation throughout life. *Eur. J. Immunol.* **23**, 3272–3277 (1993). doi: 10.1002/eji.1830231232; pmid: 8258343
32. D. R. Wesemann *et al*., Microbial colonization influences early B-lineage development in the gut lamina propria. *Nature* **501**, 112–115 (2013). doi: 10.1038/nature12496; pmid: 23965619
33. J. M. Dal Porto, A. M. Haberman, G. Kelsoe, M. J. Shlomchik, Very low affinity B cells form germinal centers, become memory B cells, and participate in secondary immune responses when higher affinity competition is reduced. *J. Exp. Med.* **195**, 1215–1221 (2002). doi: 10.1084/jem.20011550; pmid: 11994427
34. T. A. Y. Shih, E. Meffre, M. Roederer, M. C. Nussenzweig, Role of BCR affinity in T cell dependent antibody responses in vivo. *Nat. Immunol.* **3**, 570–575 (2002). doi: 10.1038/ni803; pmid: 12021782
35. J. M. Dal Porto, A. M. Haberman, M. J. Shlomchik, G. Kelsoe, Antigen drives very low affinity B cells to become plasmacytes and enter germinal centers. *J. Immunol.* **161**, 5373–5381 (1998). pmid: 9820511
36. A. T. McGuire *et al*., Antigen modification regulates competition of broad and narrow neutralizing HIV antibodies. *Science* **346**, 1380–1383 (2014). doi: 10.1126/science.1259206; pmid: 25504724
37. J. Sui *et al*., Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009). doi: 10.1038/nsmb.1566; pmid: 19234466
38. L. Morris *et al*., Isolation of a human anti-HIV gp41 membrane proximal region neutralizing antibody by antigen-specific single B cell sorting. *PLOS ONE* **6**, e23532 (2011). doi: 10.1371/journal.pone.0023532; pmid: 21980336
39. J. P. Julien *et al*., Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1477–1483 (2013). doi: 10.1126/science.1245625; pmid: 24179159
40. D. Lyumkis *et al*., Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013). doi: 10.1126/science.1245627; pmid: 24179160
41. E. Falkowska *et al*., Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* **40**, 657–668 (2014). doi: 10.1016/j.immuni.2014.04.009; pmid: 24768347
42. C. Blattner *et al*., Structural delineation of a quaternary, cleavage-dependent epitope at the gp41-gp120 interface on intact HIV-1 Env trimers. *Immunity* **40**, 669–680 (2014). doi: 10.1016/j.immuni.2014.04.008; pmid: 24768348
43. L. Scharf *et al*., Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports* **7**, 785–795 (2014). doi: 10.1016/j.celrep.2014.04.001; pmid: 24767986
44. J. Huang *et al*., Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature* **515**, 138–142 (2014). doi: 10.1038/nature13601; pmid: 25186731
45. T. Muster *et al*., Cross-neutralizing activity against divergent human immunodeficiency virus type 1 isolates induced by the gp41 sequence ELDKWAS. *J. Virol.* **68**, 4031–4034 (1994). pmid: 7514684
46. J. Huang *et al*., Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012). doi: 10.1038/nature11544; pmid: 23151583
47. P. A. Goepfert *et al*., Specificity and 6-month durability of immune responses induced by DNA and recombinant modified vaccinia Ankara vaccines expressing HIV-1 virus-like particles. *J. Infect. Dis.* **210**, 99–110 (2014). doi: 10.1093/infdis/jiu003; pmid: 24403557
48. Q. Li *et al*., Live simian immunodeficiency virus vaccine correlate of protection: Local antibody production and concentration on the path of virus entry. *J. Immunol.* **193**, 3113–3125 (2014). doi: 10.4049/jimmunol.1400820; pmid: 25135832
49. H. L. Robinson, Non-neutralizing antibodies in prevention of HIV infection. *Expert Opin. Biol. Ther.* **13**, 197–207 (2013). doi: 10.1517/14712598.2012.743527; pmid: 23130709
50. D. R. Burton *et al*., Limited or no protection by weakly or nonneutralizing antibodies against vaginal SHIV challenge of macaques compared with a strongly neutralizing antibody. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11181–11186 (2011). doi: 10.1073/pnas.1103012108; pmid: 21690411
51. S. Kwa *et al*., CD40L-adjuvanted DNA/modified vaccinia virus Ankara simian immunodeficiency virus (SIV) vaccine enhances protection against neutralization-resistant mucosal SIV infection. *J. Virol.* **89**, 4690–4695 (2015). doi: 10.1128/JVI.03527-14; pmid: 25653428
52. B. F. Haynes, L. Verkoczy, AIDS/HIV. Host controls of HIV neutralizing antibodies. *Science* **344**, 588–589 (2014). doi: 10.1126/science.1254990; pmid: 24812389
53. G. D. Tomaras, B. F. Haynes, Lessons from babies: Inducing HIV-1 broadly neutralizing antibodies. *Nat. Med.* **20**, 583–585 (2014). doi: 10.1038/nm.3598; pmid: 24901564
54. L. Goo, V. Chohan, R. Nduati, J. Overbaugh, Early development of broadly neutralizing antibodies in HIV-1-infected infants. *Nat. Med.* **20**, 655–658 (2014). doi: 10.1038/nm.3565; pmid: 24859529
55. E. S. Gray *et al*., Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. *J. Virol.* **85**, 7719–7729 (2011). doi: 10.1128/JVI.00563-11; pmid: 21613396
56. H. X. Liao *et al*., Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* **38**, 176–186 (2013). doi: 10.1016/j.immuni.2012.11.011; pmid: 23313589
57. J. Friedman *et al*., Isolation of HIV-1-neutralizing mucosal monoclonal antibodies from human colostrum. *PLOS ONE* **7**, e37648 (2012). doi: 10.1371/journal.pone.0037648; pmid: 22624058
58. S. M. Alam *et al*., Antigenicity and immunogenicity of RV144 vaccine AIDSVAX clade E envelope immunogen is enhanced by a gp120 N-terminal deletion. *J. Virol.* **87**, 1554–1568 (2013). doi: 10.1128/JVI.00718-12; pmid: 23175357
59. D. C. Montefiori *et al*., Magnitude and breadth of the neutralizing antibody response in the RV144 and Vax003 HIV-1 vaccine efficacy trials. *J. Infect. Dis.* **206**, 431–441 (2012). doi: 10.1093/infdis/jis367; pmid: 22634875
60. E. S. Gray *et al*., Broad neutralization of human immunodeficiency virus type 1 mediated by plasma antibodies against the gp41 membrane proximal external region. *J. Virol.* **83**, 11265–11274 (2009). doi: 10.1128/JVI.01359-09; pmid: 19692477
61. J. Pollara *et al*., High-throughput quantitative analysis of HIV-1 and SIV-specific ADCC-mediating antibody responses. *Cytometry A* **79A**, 603–612 (2011). doi: 10.1002/cyto.a.21084; pmid: 21735545
62. M. Bonsignori *et al*., Antibody-dependent cellular cytotoxicity-mediating antibodies from an HIV-1 vaccine efficacy trial target multiple epitopes and preferentially use the VH1 gene family. *J. Virol.* **86**, 11521–11532 (2012). doi: 10.1128/JVI.01023-12 pmid: 22896626
63. P. Liu *et al*., Infectious virion capture by HIV-1 gp120-specific IgG from RV144 vaccinees. *J. Virol.* **87**, 7828–7836 (2013). doi: 10.1128/JVI.02737-12; pmid: 23658446
64. K. Smith *et al*., Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat. Protoc.* **4**, 372–384 (2009). doi: 10.1038/nprot.2009.3; pmid: 19247287
65. B. Ewing, P. Green, Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998). doi: 10.1101/gr.8.3.175; pmid: 9521922
66. B. Ewing, L. Hillier, M. C. Wendl, P. Green, Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998). doi: 10.1101/gr.8.3.175; pmid: 9521921
67. T. B. Kepler *et al*., Chiropteran types I and II interferon genes inferred from genome sequencing traces by a statistical gene-family assembler. *BMC Genomics* **11**, 444 (2010). doi: 10.1186/1471-2164-11-444; pmid: 20663124
68. T. F. Smith, M. S. Waterman, Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981). doi: 10.1016/0022-2836(81)90087-5; pmid: 7265238
69. J. M. Volpe, L. G. Cowell, T. B. Kepler, SoDA: Implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics* **22**, 438–444 (2006). doi: 10.1093/bioinformatics/btk004; pmid: 16357034
70. M. Bonsignori *et al*., Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85**, 9998–10009 (2011). doi: 10.1128/JVI.05045-11; pmid: 21795340
71. H. X. Liao *et al*., Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013). doi: 10.1038/nature12053; pmid: 23552890
72. T. B. Kepler *et al*., Reconstructing a B-cell clonal lineage. II. Mutation, selection, and affinity maturation. *Front. Immunol.* **5**, 170 (2014). doi: 10.3389/fimmu.2014.00170; pmid: 24795717

ACKNOWLEDGMENTS

The authors thank G. Kelsoe for manuscript review, L. Armand for production of fluorophore-labeled reagents, and C. Stolarchuk, S. Stewart, A. Wang, R. Duffy, A. Deal, J. Eudailey, T. Von Holle, J. Alin, L. Oliver, F. Jaeger, and S. Arora for technical assistance. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Research materials used in this study are available from Duke University upon request and subsequent execution of an appropriate materials transfer agreement. Supported by NIH, NIAID UM-1 grant Center for HIV/AIDS Vaccine Immunology-Immunogen Discovery (CHAVI-ID; UM1 AI100645), NIH NIAID Duke University Center for AIDS Research (CFAR; P30-AI-64518), the NIH NIAID HVTN Laboratory Center UM1AI068618, and the intramural research program of the Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH. W.B.W. designed and performed experiments, analyzed data, and co-wrote the paper; H.-X.L., K.J., M.A.M., D.J.M., and J.F.W. performed isolation of Abs, reviewed data, and edited the paper; T.B.K., A.H., K.W., K.S., and A.M.T. performed computational analysis of Ab sequences; F.G., R.Z., and H.S. sequenced Abs and performed NGS; S.M.A., P.L., M.Z.T., K.E.S., X.S., A.F., K.E.L., R.P., and J.-S.Y. performed Ab binding and functional assays; J.P. and G.F. performed ADCC assays; N.V., D.G., and P.G. performed statistical analysis; D.C.M. performed neutralization assays; M.E.S., S.H., S.K., N.G., M.J.M., J.R.M., R.A.K., L.C, G.J.N., C.M., G.C., J.M., M.K., B.S.G., and L.R.B. were members of the VRC or HVTN teams that carried out the clinical trials; G.D.T. analyzed data, designed experiments, reviewed data, and edited the paper; and B.F.H. conceived and designed the study, performed ANA analysis, reviewed all data, and cowrote the paper. GenBank accession numbers for sequences of the 221 Abs isolated via flow cytometry memory B cell single-cell sorting: immunoglobulin heavy chains, KT304331–KT304551; immunoglobulin light chains, KT304552–KT304772.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/aab1253/suppl/DC1
Figs. S1 to S12
Tables S1 to S25

15 March 2015; accepted 9 July 2015
Published online 30 July 2015
10.1126/science.aab1253

RESEARCH ARTICLES

PALEOCEANOGRAPHY

A warm and poorly ventilated deep Arctic Mediterranean during the last glacial period

D. J. R. Thornalley,[1,2]* H. A. Bauch,[3] G. Gebbie,[1] W. Guo,[1] M. Ziegler,[4] S. M. Bernasconi,[4] S. Barker,[5] L. C. Skinner,[6] J. Yu[7]

Changes in the formation of dense water in the Arctic Ocean and Nordic Seas [the "Arctic Mediterranean" (AM)] probably contributed to the altered climate of the last glacial period. We examined past changes in AM circulation by reconstructing radiocarbon ventilation ages of the deep Nordic Seas over the past 30,000 years. Our results show that the glacial deep AM was extremely poorly ventilated (ventilation ages of up to 10,000 years). Subsequent episodic overflow of aged water into the mid-depth North Atlantic occurred during deglaciation. Proxy data also suggest that the deep glacial AM was ~2° to 3°C warmer than modern temperatures; deglacial mixing of the deep AM with the upper ocean thus potentially contributed to the melting of sea ice, icebergs, and terminal ice-sheet margins.

The Atlantic Meridional Overturning Circulation (AMOC) plays an important role in Earth's climate, because it redistributes ocean heat and helps control the storage of carbon in the deep ocean. The primary Northern Hemisphere sources of dense water supplied to the AMOC are produced in the Arctic Mediterranean (AM) (*1*). Warm surface waters from the Atlantic flow northward and circulate around the AM via several different pathways, gradually cooling (thereby releasing heat to the atmosphere) and becoming denser. Much of this water-mass transformation is thought to occur in the Nordic Seas via intermediate and deep open-ocean convection, with a smaller contribution from the Arctic Ocean involving the addition of dense waters from brine-enhanced shelf water production (*1*, *2*). The dense water produced by these processes overflows the Greenland-Scotland Ridge, ultimately forming lower North Atlantic Deep Water (NADW) as part of the deep southward return flow of the AMOC.

[1]Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA 02543, USA. [2]Department of Geography, University College London, London WC1E 6BT, UK. [3]Academy of Sciences, Humanities and Literature, Mainz, and GEOMAR Helmholtz Centre for Ocean Research, 24148 Kiel, Germany. [4]Eidgenössische Technische Hochschule Zürich, 8093 Zurich, Switzerland. [5]School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3XQ, UK. [6]Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK. [7]Research School of Earth Sciences, Australian National University, Acton, ACT 0200, Australia.
*Corresponding author. E-mail: d.thornalley@cantab.net

Because of the northward heat transfer associated with the flow of warm surface water to convection sites, changes in deep-water formation in the North Atlantic and Nordic Seas are thought to be associated with the altered climate of the Last Glacial Maximum (LGM) and the abrupt climate events of the last deglaciation [~19 to 7 thousand years ago (ka)], such as the Northern Hemisphere cold intervals Heinrich Stadial 1 (HS1) and the Younger Dryas (YD) (*3*–*5*), which affected global climate (*6*, *7*). In this study, we investigated circulation changes over the past 30 ka in the deep Norwegian Sea (and by inference, the broader AM) by reconstructing radiocarbon (^{14}C) ventilation ages and deep-ocean temperatures. Our results revealed an absence of deep convection within the AM throughout much of the last glacial period and the subsequent deglaciation; they instead suggest the presence of a relatively warm and extremely poorly ventilated water mass in the glacial deep AM that subsequently overflowed southward into the North Atlantic during the deglaciation.

North Atlantic radiocarbon reconstructions

Several studies have used seawater radiocarbon ratios (Δ^{14}C) as a proxy for investigating past changes in the circulation of the North Atlantic (*8*–*12*). In the modern high-latitude North Atlantic, deep convection in the Nordic and Labrador Seas quickly transfers surface waters that have equilibrated with the atmosphere to the deep ocean, resulting in a minimal surface-to-deep gradient in ^{14}C age (~100 years) (*13*) and well-ventilated deep water in the North Atlantic.

Deglacial Δ^{14}C reconstructions from the subtropical North Atlantic are consistent with the established view that there was shoaling of the AMOC [involving a switch from NADW formation to Glacial North Atlantic Intermediate Water (GNAIW) formation] and a northward incursion into the deep North Atlantic of ^{14}C-depleted southern-sourced water (SSW) (*8*, *11*) during the last glacial period, HS1, and the YD. However,

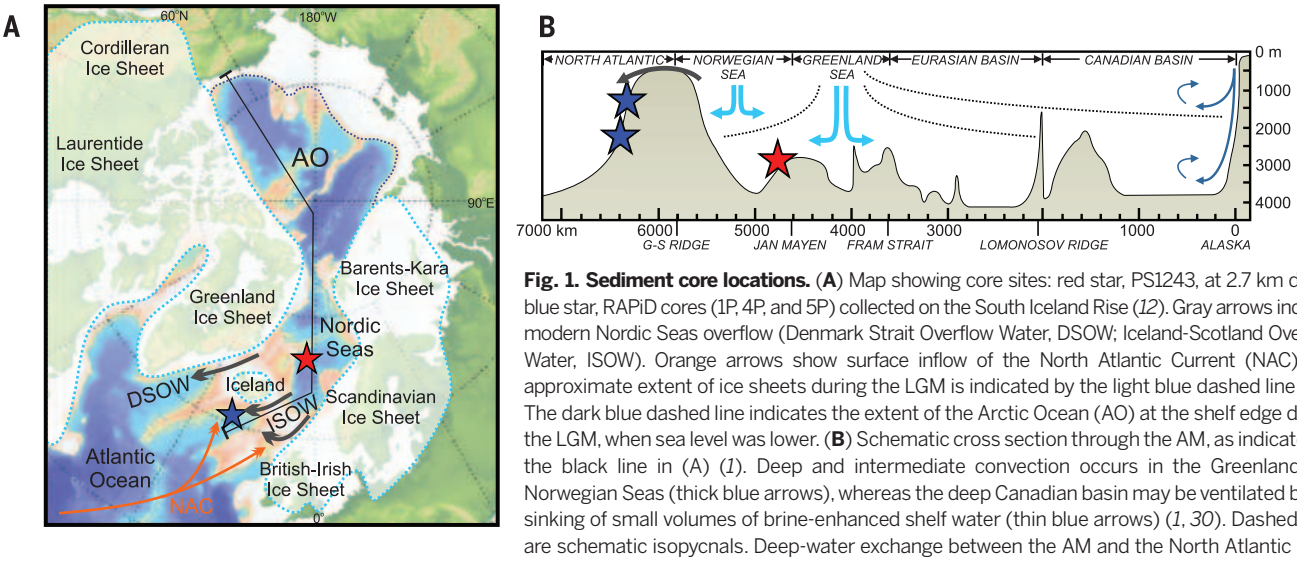


Fig. 1. Sediment core locations. (**A**) Map showing core sites: red star, PS1243, at 2.7 km depth; blue star, RAPiD cores (1P, 4P, and 5P) collected on the South Iceland Rise (*12*). Gray arrows indicate modern Nordic Seas overflow (Denmark Strait Overflow Water, DSOW; Iceland-Scotland Overflow Water, ISOW). Orange arrows show surface inflow of the North Atlantic Current (NAC). The approximate extent of ice sheets during the LGM is indicated by the light blue dashed line (*26*). The dark blue dashed line indicates the extent of the Arctic Ocean (AO) at the shelf edge during the LGM, when sea level was lower. (**B**) Schematic cross section through the AM, as indicated by the black line in (A) (*1*). Deep and intermediate convection occurs in the Greenland and Norwegian Seas (thick blue arrows), whereas the deep Canadian basin may be ventilated by the sinking of small volumes of brine-enhanced shelf water (thin blue arrows) (*1*, *30*). Dashed lines are schematic isopycnals. Deep-water exchange between the AM and the North Atlantic (gray arrow) is restricted by the Greenland-Scotland Ridge, with a depth of ~400 to 800 m.

these data also reveal that large fluctuations in $\Delta^{14}C$ (shifts of up to ~1000 ^{14}C years within ~100 calendar years) occurred in the mid-depth North Atlantic during HS1 and the YD, suggesting a complex history of circulation (*8*). Recently, these fluctuations in mid-depth $\Delta^{14}C$, alongside temperature proxy data, have been cited as evidence that the AMOC switched from a glacial to an interglacial mode of circulation through the release of heat from warm deep water (*14*).

$\Delta^{14}C$ reconstructions from the mid-depth (1.2 to 2.3 km) subpolar Northeast Atlantic, south of Iceland (at the South Iceland Rise), have revealed the presence of an extremely poorly ventilated water mass during cold intervals of the last deglaciation, with ^{14}C ventilation ages in excess of 5000 years (*12*). In addition, rapid and large fluctuations in ventilation ages occurred during HS1 and the YD. These fluctuations are similar to the variability that has been reconstructed in the subtropical Northwest Atlantic, although the shifts south of Iceland are up to four times larger in amplitude. The source of the poorly ventilated water south of Iceland was initially interpreted as Antarctic Intermediate Water (AAIW) (*12*), but more recent studies have shown that AAIW in the deglacial Atlantic was not as poorly ventilated as the water south of Iceland (*9*, *15*, *16*). Available evidence also suggests that deep SSW was not sufficiently depleted in ^{14}C to explain the South Iceland Rise data (*8*–*10*), nor can it explain the observed distinct relationship between $\delta^{13}C$ and $\Delta^{14}C$ (*12*). An alternative proposed source is the AM, although thus far there have been no ^{14}C ventilation data from the deep AM to test this hypothesis. Reconstructions of ^{14}C ventilation ages from the shallow (~700 m) Iceland Sea during the last glacial period reveal benthic-atmosphere ventilation ages of ~500 years, indicating well-ventilated intermediate-depth water in the Nordic Seas that may have contributed to GNAIW formation (*3*, *17*).

This study was therefore motivated by two aims. First, we sought to increase the understanding of AM circulation and to investigate whether there was continued deep convection in the Nordic Seas (or the AM as a whole) during the last glacial period. It is important to better constrain past circulation changes in the AM because (i) the amount of exchange between the surface and deep AM alters the properties of the dense water it exports (*1*, *2*); (ii) regional and global climate are directly influenced by the northward heat transport associated with the inflow of warm surface waters feeding the high-latitude dense water formation sites (*2*); and (iii) ocean circulation changes have the potential to affect other components of the climate system, such as sea-ice extent and adjacent ice sheets. We also wished to investigate the cause of the mid-depth radiocarbon anomalies in the South Iceland Rise data. The present lack of a viable explanation for these data suggests a knowledge gap in our understanding of deglacial ocean circulation. Moreover, a more complete interpretation of deglacial variability in mid-depth North Atlantic ^{14}C ventilation ages—which, for example, have been examined in recent studies of data from the New England seamounts (*14*, *18*)—first requires us to constrain the various end-member water masses and the mechanisms by which they formed.

Ventilation changes in the AM

We obtained new benthic and planktic radiocarbon measurements from a marine sediment core collected in the Norwegian Sea (core PS1243, 2.7 km depth) (Fig. 1), which we selected because of its well-defined stratigraphy (*19*). The age model was slightly modified from its original planktic ^{14}C–based chronology by tuning planktic $\delta^{18}O$ to nearby cores that had been placed on a Greenland ice core–based age model (Fig. 2 and supplementary materials). Core sites from within the central Arctic Ocean were not chosen because of the low abundance of benthic foraminifera; in addition, the extremely low glacial sedimentation rate (*20*) would have increased the uncertainty in the stratigraphic age of samples and hence in the reconstruction of $\Delta^{14}C$.

In contrast to the well-ventilated deep waters and the small benthic-planktic (B-P) ^{14}C offset (~100 years) of the Holocene and the present day, Table 1 and Fig. 2 show that much of the late glacial and deglacial Norwegian Sea was characterized by extremely poorly ventilated deep waters, with benthic ventilation ages of ~7000 to 10,000 years (and B-P ^{14}C offsets of a similar magnitude). Moreover, these extremely old ventilation ages were only associated with a modest decrease in benthic $\delta^{13}C$ values. The coupling of relatively high $\delta^{13}C$ values (~0.8 to 1.4‰) with extremely old ^{14}C ventilation ages suggests that the aging of deep waters was not accompanied by substantial remineralization of organic matter at depth, probably reflecting low surface productivity (Fig. 3 and supplementary materials). These data further suggest that the deep Norwegian Sea was the likely source of poorly ventilated water south of Iceland during the deglaciation (*12*), and the slightly lower $\delta^{13}C$ values reconstructed for deglacial cold intervals south of Iceland can be explained by mixing between the Nordic overflow and low-$\delta^{13}C$ SSW (Fig. 3). The observation of a poorly ventilated yet high-$\delta^{13}C$ end member highlights the complexity of interpreting benthic $\delta^{13}C$ in the subpolar Northeast Atlantic, because it cannot be interpreted as a simple two–end-member mixing scenario (i.e., SSW versus GNAIW or NADW). Our results demonstrate that during the deglaciation, a Nordic Seas overflow of extremely poorly ventilated water with $\delta^{13}C$ values of ~0.8 to 1.4‰ took place; in addition, the chemically distinct contribution to GNAIW during the LGM identified in previous research (*21*) must have been sourced from the intermediate (not deep) Nordic Seas. The high $\delta^{13}C$ values of the *Cibicidoides wuellerstorfi* at this site and the benthic foraminifera samples used

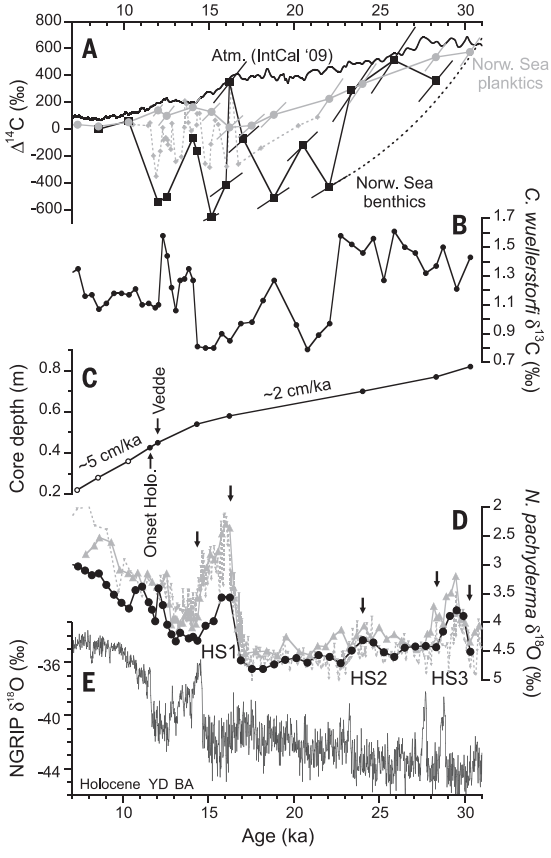


Fig. 2. Deep Norwegian Sea radiocarbon reconstructions. (**A**) $\Delta^{14}C$ from core PS1243, collected in the deep Norwegian Sea (black squares, benthic species; gray circles, planktic species), and from the South Iceland Rise (1.2 to 2.3 km depth; benthics, small gray diamonds) (*12*), shown with the IntCal09 radiocarbon age calibration curve (*40*). "Projection age" extrapolation for the sample at 23 ka is shown by the black dashed line. (**B**) *C. wuellerstorfi* $\delta^{13}C$ measurements from core PS1243 (*19*). (**C**) Age model for core PS1243. White circles are tie points based on planktic ^{14}C using the modern reservoir age of 400 years. Black circles are stratigraphic tie points based on the correlation of *Neogloboquadrina pachyderma* (s) $\delta^{18}O$ data [shown in (D) by black arrows], the occurrence of the Vedde Ash, and an abrupt decrease in the percentage of *N. pachyderma* (s) (not shown) at the onset of the Holocene. (**D**) Correlation of *N. pachyderma* (s) $\delta^{18}O$ measurements between core PS1243 (black) (*19*) and Norwegian Sea cores ENAM93-21 (solid gray) and MD952010 (dashed gray). The cores have been placed on the Greenland Ice Core Chronology 2005 age scale of the North Greenland Ice Core Project (NGRIP) (**E**) on the basis of their magnetic susceptibility (*5*, *41*).

for $\Delta^{14}C$ (table S1) rule out low-$\delta^{13}C$ sources of ^{14}C-depleted carbon, such as methane hydrates, mantle carbon, or remineralized sedimentary organic carbon.

Our records indicate enhanced ventilation at the onset of the Holocene and the Bølling-Allerød (BA) periods, although additional measurements are required for confirmation (see the supplementary materials); they also point to an event at ~16 ka that is in agreement with measurements from south of Iceland. Support for the existence of a brief (multicentennial-scale or less) deep-water formation event at ~16 ka, with a benthic-atmosphere offset of ~300 years recorded both in the deep Norwegian Sea (at 2.7 km depth) and south of Iceland (at 2.3 km depth) (*12*), can also be found in high-resolution records of planktic foraminifer faunal assemblages in the subpolar Northeast Atlantic (fig. S3 and supplementary materials). These records reveal a strong surface warming farther to the south, possibly caused by an increased northward flow of surface waters feeding the deep convection site, the precise location of which is uncertain.

Table 1. Benthic ^{14}C data from core PS1243. Intervals for benthic dates were selected based on local abundance peaks where available (supplementary materials). For some samples, a robust B-P age is not provided, because there was no planktic ^{14}C date at the exact corresponding core depth. Foraminifera species used in this analysis include *C. wuellerstorfi* (*Cw*), *P. depressa* (*Pyrgo*), and Miliolida species.

Time interval	Depth (m)	Calendar age (ka)	Species	^{14}C age (years)	Error (years)	B-P offset (years)	Error (years)	Benthic-atmosphere offset (years)
Holocene	0.28	8.52	*Cw*	8290	35	170	100	570
	0.28	8.52	*Pyrgo*	8090	75	−30	110	370
	0.36	10.29	*Cw*	9530	40	−70	100	340
YD (Vedde)	0.45	12.03	*Cw*	18,000	100	7370	140	7760
YD	0.47	12.52	*Pyrgo*	17,800	150	6400	160	7310
BA	0.53	14.05	*Pyrgo/Cw*	14,250	65	1810	91	2030
	0.54	14.30	Miliolida	15,350	100	N/A		2930
HS1	0.56	15.15	*Pyrgo*	23,200	250	9450	290	10,400
	0.575	16.00	*Cw*	19,850	130	N/A		6710
	0.58	16.20	*Pyrgo*	13,350	60	−2300	140	172
	0.595	17.00	*Pyrgo/Cw*	17,100	95	N/A		3110
Late glacial	0.62	18.80	Miliolida	24,000	290	6550	350	8370
	0.645	20.50	Miliolida	20,900	160	N/A		3530
	0.67	22.05	Miliolida	26,000	270	6210	340	7640
Early MIS 2	0.69	23.35	*Pyrgo*	20,700	660	N/A		1220
	0.73	25.84	*Pyrgo*	21,800	220	N/A		270
	0.77	28.30	Miliolida	25,000	240	950	350	1520

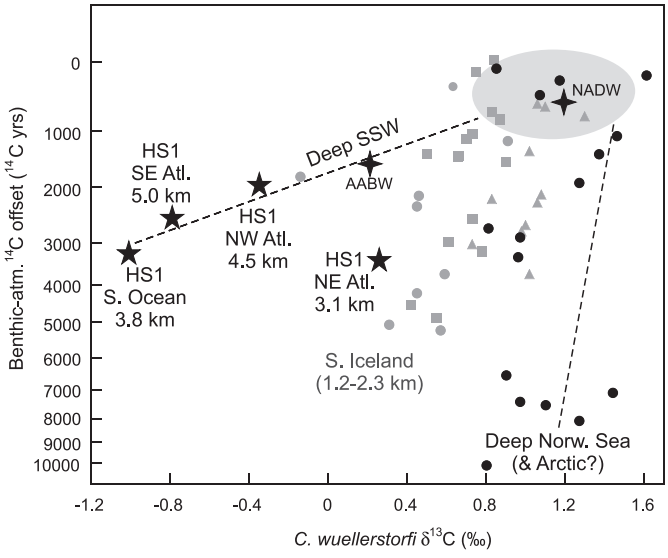


Fig. 3. Cross-plot of *C. wuellerstorfi* $\delta^{13}C$ and benthic-atmosphere ^{14}C ventilation ages (logarithmic scale), modified from (*12*). Black circles are new data from the deep Norwegian Sea; gray shapes are data from RAPiD cores previously collected on the South Iceland Rise (circles, core 5P; squares, core 4P; triangles, core 1P). Five-pointed stars are published estimates for HS1, and four-pointed stars are estimates for modern water masses.

Isolation of the deep central Arctic Ocean

The magnitude and rapidity of the shift from well-ventilated to poorly ventilated water at ~23 ka cannot be explained by in situ aging of deep water at the Norwegian Sea core site; therefore, we infer the incursion of a pre-aged water mass. The most likely candidate for this is shoaling of isolated water in the deep central Arctic Ocean (supplementary materials), which is connected to the Nordic Seas via the Fram Strait and which has a volume of ~1.2×10^7 km^3—approximately four times greater than the deep (>1 km) Nordic Seas. Because the Arctic Ocean (and the AM as a whole) is a semi-enclosed basin that was further restricted during the last glacial period by the closure of the Bering Strait, it was susceptible to isolation and the development of poorly ventilated deep water.

Previous studies have suggested that there was thick ice cover across much of the Arctic Ocean during the LGM, leading to minimal or no sedimentation in the central Arctic of either terrigenous or biogenic material (*20*, *23*). With no appreciable surrounding continental-shelf area, there was probably also only a minor contribution to the deep ocean from brine-enhanced shelf water (*24*). These conditions would favor the development of poorly ventilated bottom water with relatively high $\delta^{13}C$. Benthic $\delta^{13}C$ data from the deep central Arctic before the last glacial period [Marine Isotope Stage (MIS) 3] suggest values of up to 1.7 to 1.8‰ (*25*), which, with a small amount of organic carbon remineralization throughout the glacial period (MIS 2), would have resulted in the observed glacial and deglacial values in the deep Nordic Seas of 0.8 to 1.4‰. Reduced surface productivity across much of the glacial AM and organic carbon export to the North Atlantic by vigorous intermediate-depth circulation presumably helped to prevent the development of anoxia in the deep AM (supplementary materials).

Projecting the benthic $\Delta^{14}C$ onto the atmospheric $\Delta^{14}C$ curve (*11*) (Fig. 2) for the data point at ~23 ka (i.e., the first occurrence of old water in data from core PS1243), using a surface reservoir age of 400 years, suggests that the deep Arctic Ocean may have become isolated at ~30 ka (i.e., close to the onset of MIS 2). At this time, sea-level records indicate the final rapid growth of continental ice sheets to their full glacial extent (*26*) and hence the loss of shelf seas, reducing vertical mixing (*27*) and brine-enhanced shelf water production (*24*). However, this method does not consider any mixing or entrainment with younger waters. Therefore, if such entrainment occurred, it is possible that parts of the AM were older than our recorded ages and that the deep central AM was isolated earlier than ~30 ka. Given its complex bathymetry, it is plausible that there were numerous distinct water masses residing within the deep AM during the

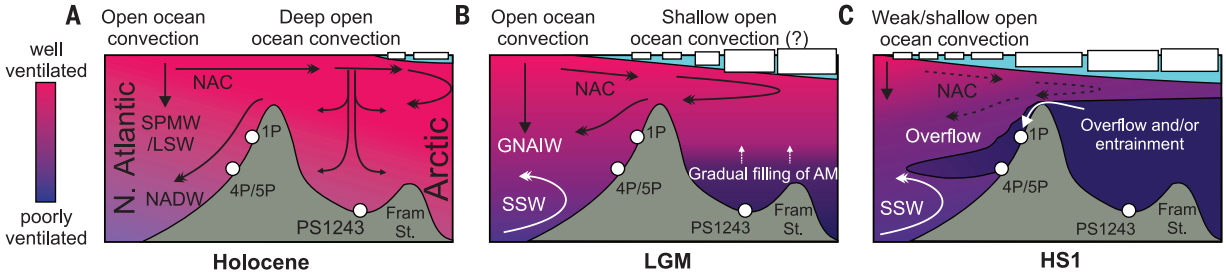


Fig. 4. Cartoon of hypothesized changes in the circulation and ventilation of the Arctic Ocean, Nordic Seas, and Northeast Atlantic. North is to the right. Color shading indicates ^{14}C ventilation. The pale blue layer indicates the varying extent of surface fresh water and sea ice or icebergs (white rectangles). Core locations are indicated by white circles. (**A**) Surface waters flow around the Nordic Seas and Arctic Ocean, losing buoyancy en route (SPMW, Subpolar Mode Water; LSW, Labrador Sea Water); ventilation of the intermediate and deep AM occurs mainly by convection in the Nordic Seas. (**B**) Reduction of deep convection and accumulation of poorly ventilated dense water in the deep AM, resulting in a sharp radiocarbon front (supplementary materials). (**C**) Entrainment and/or overflow of aged AM water into the North Atlantic during HS1.

last glacial period, with various ventilation ages and nutrient chemistry.

Circulation changes and rates

Deep open-ocean convection in the modern Nordic Seas is variable and subject to perturbation (*28*); therefore, it seems probable that during the late glacial and deglacial periods, enhanced surface freshening caused the weakening and shoaling of open-ocean convection in the Nordic Seas. This would mean that the deep AM was only ventilated by limited deep-water formation, probably in coastal polynyas. The timing of the switch to poorly ventilated waters at ~23 ka in the deep Norwegian Sea may have been the result of a marked weakening or shoaling of deep convection in response to surface freshening from melting icebergs (*19*). Alternatively, it may simply have been caused by the slow accumulation of bottom water in the AM that was denser than the products of open-ocean convection in the Nordic Seas. Thus, the rapid switch to older benthic ventilation ages in core PS1243 at ~23 ka probably records the upward migration of a sharp vertical gradient in radiocarbon ventilation age, separating the aged AM bottom waters from the overlying, better-ventilated products of open-ocean convection (supplementary materials). This is somewhat analogous to the model invoked for the glacial Atlantic Ocean, with dense SSW, formed partly by brine rejection processes around Antarctica, being overlain by well-ventilated GNAIW, the product of open-ocean convection south of Iceland (*29*). The mechanism by which the aged water mass was subsequently transported into the North Atlantic is uncertain, but it must have involved either (i) the continued slow accumulation of aged water in the AM and its eventual overflow (Fig. 4) or (ii) a more rapid process of entrainment and displacement by overturning in the Nordic Seas during the deglaciation, possibly related to sea-ice formation and brine rejection as orbitally induced insolation changes promoted seasonal melting and refreezing of sea ice. Investigating these mechanisms will require a depth transect of cores to reconstruct water-column $\Delta^{14}C$ profiles at time intervals throughout the glacial and early deglacial periods.

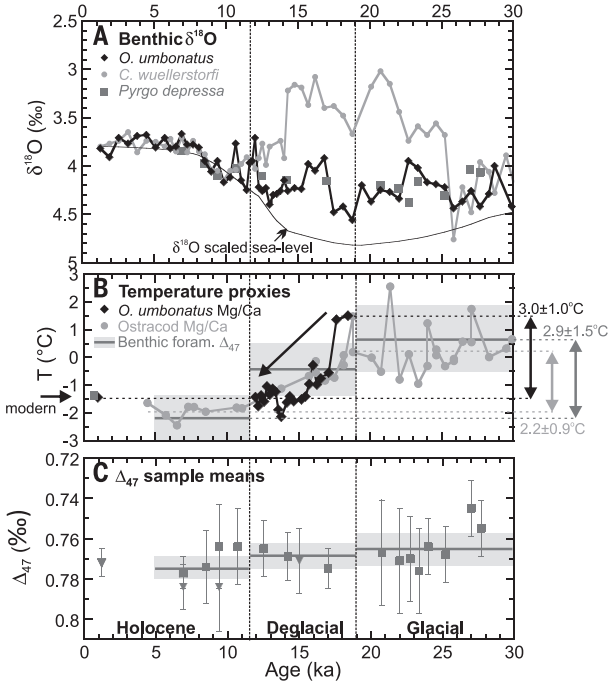


Fig. 5. Deep Norwegian Sea temperature reconstructions. (**A**) Benthic $\delta^{18}O$ data from core PS1243 (*19*). Measurements from *Pyrgo depressa* (gray squares; vital effect, −0.9‰), *C. wuellerstorfi* (gray circles), and *Oridosarlis umbonatus* (black diamonds; vital effect, −0.28‰) are shown with global sea level scaled to a 1‰ whole-ocean $\delta^{18}O$ change. The low $\delta^{18}O$ values of *C. wuellerstorfi* are discussed in the supplementary materials. (**B**) Glacial-to-Holocene temperature change in the deep Norwegian Sea. Shown are new clumped isotope data (Δ_{47}, horizontal gray line; gray square, core top) from core PS1243, published ostracod Mg/Ca data (gray circles) from core PS1243 (*35*), and new benthic foraminifera Mg/Ca data (black diamonds) from core MD992276 (same site as PS1243) (supplementary materials). Dashed lines indicate the Holocene (lower) and glacial (upper) averages used for each proxy to calculate the glacial-to-Holocene temperature change (double-headed arrows, with ±1 SE); the Δ_{47} data set is also shown with ±1σ errors for each time interval (shading). The black arrow in the center highlights the deglacial release of heat from the deep Norwegian Sea. (**C**) Sample-mean Δ_{47} measurements from *P. depressa* (squares) and *C. wuellerstorfi* (triangles) in core PS1243, with averages (lines) and ±2 SE shading for the Holocene, deglacial, and glacial intervals. Core-top foraminifera Mg/Ca and Δ_{47} data points are from (*42*) and (*43*), respectively.

If the deep ventilation age of 10,000 years is interpreted as a water-mass residence time of ~10,000 years in the deep AM, then we can infer that the replenishment rate of deep water must have been limited to ~0.05 sverdrup [1 sverdrup (Sv) = 10^6 m^3 s^{-1}], presumably by processes such as brine-enhanced dense water production in coastal polynyas. This rate can be compared with estimates of ~0.1 Sv for the modern contribution of brine-enhanced shelf-slope and entrained water to the deep Canadian Basin (*30*). Employing estimates of vertical diffusivity from the modern Arctic in a simple advective-diffusive model of the deep Arctic demonstrates that ventilation ages similar to those that we reconstructed in the glacial deep AM (~10,000 years) are achievable, as long as bottom-water formation rates remain low (supplementary materials).

Recent work has suggested that there has been a persistent export of ^{231}Pa from the deep central Arctic Ocean over the past 35 ka (*31*). The Fram Strait and Nordic Seas are a likely sink for this ^{231}Pa because of their higher particle fluxes (*32*), enabling scavenging of dissolved ^{231}Pa from the water column. To reconcile this with our results, we must infer that there was sufficient recirculation

of the poorly ventilated deep water within the AM to enable the export of dissolved ^{231}Pa to the Fram Strait and Nordic Seas. However, because uncertainties also remain regarding the influence of reduced particle rain and boundary scavenging on the Arctic ^{231}Pa budget during glacial periods, these factors may also be called upon to explain the ^{231}Pa deficit in the glacial Arctic Ocean (*31*).

Implications

Using simple mass-balance calculations that include the possible volume of the poorly ventilated reservoir (i.e., the volume of the deep AM), limitations on its renewal rate (~0.05 Sv), and the observations from south of Iceland indicating overflow for a total of ~5000 years [i.e., the combined duration of HS1, the YD, and the Intra-Allerød Cold Period, when highly ^{14}C-depleted water was present in the overflow region (*12*)], we conclude that the overflow of poorly ventilated waters from the AM into the Northeast Atlantic must have been relatively weak and would not have exceeded, on average, ~0.1 Sv. Future modeling work, aided by additional proxy reconstructions, should investigate whether the mid-depth Δ^{14}C variability reconstructed for the subtropical Northwest Atlantic (*8*) can be attributed to a weak yet highly ^{14}C-depleted Nordic Seas overflow (*18*) and how this signal propagated throughout the North Atlantic under reduced AMOC conditions. Possibly the AM was the source for the low-Δ^{14}C event at 15.6 ka that was recorded in corals at the New England seamounts (*14*, *18*). However, the reconstructed temperatures of the deglacial AM are colder than those reconstructed at the New England seamounts for the 15.6 ka event (~0° to 1°C versus 3° to 4°C); therefore, substantial mixing with a warmer water mass must also have occurred.

The evidence for poorly ventilated conditions in the deep Nordic Seas also enables us to conclude that there was no significant contribution from the deep Nordic Seas (and by inference, from the deep Arctic Ocean) to the formation of GNAIW. This contrasts with suggestions by earlier workers (*33*) that relied solely on benthic δ^{13}C. Given faunal evidence for a persistent Atlantic inflow to the Nordic Seas (*34*) and possibly to the Arctic as a subsurface layer (*35*), shallow to intermediate overturning in the Nordic Seas probably persisted throughout the LGM, contributing to GNAIW, as suggested by previous studies (*3*, *17*, *19*, *21*, *36*). Also, while the deep AM remained isolated from the overlying, better-ventilated upper ocean, any surface inputs or perturbations to the AM, such as meltwater events, would not have mixed throughout the entire basin; rather, they would have largely been confined to the upper ocean.

Warm temperatures in the glacial deep AM

Because of deep convection, the modern deep Nordic Seas have a temperature of approximately −1° to −1.5°C. Published temperature proxy data, obtained using ostracod Mg/Ca ratios from core PS1243, indicate that temperatures of the deep Norwegian Sea during the last glacial period were ~2° to 3°C warmer than they have been during the Holocene (*35*). In agreement with these estimates, newly obtained clumped isotope (Δ_{47}) and Mg/Ca temperature proxy data from benthic foraminifera in core PS1243 and neighboring core MD992276 indicate that the LGM was warmer than the Holocene by 2.9 ± 1.5°C (Δ_{47}) or 3.0 ± 1.0°C (Mg/Ca) (Fig. 5 and supplementary materials). An intermediate-to-deep AM during MIS 3 that was 2° to 4°C warmer than modern temperatures, caused by a deep inflow of Atlantic water, has been inferred (*24*, *35*). During MIS 2, when we infer an isolated deep AM, it is likely that geothermal heating contributed to warming in the deep AM. The contribution of geothermal heating to the modern Arctic Ocean is 40 to 60 mW m^{-2} (*37*); assuming that all the geothermal heat remained in a bottom-water layer 2000 m thick, the bottom water could have warmed by ~2°C over 10,000 years.

Because of the thermobaric effect, the development of warm and presumably relatively salty deep water, when overlain by colder and fresher water, provides a source of potential energy that can help drive ocean mixing and overturning (*38*). Overturning of the deep Nordic Seas (or the Arctic) would allow the release of heat previously stored in the deep ocean, which could promote the melting of sea ice and destabilize marine-terminating ice sheets, such as the Barents Sea ice shelf, thereby contributing to the deglaciation of the region. Ostracod Mg/Ca data (*35*) and our new benthic foraminifera Mg/Ca data from the Norwegian Sea suggest that much of the heat stored in the deep Nordic Seas was released over the interval from ~18 to 15 ka (i.e., during the early deglacial period, centered around HS1) (Fig. 5). If deep AM waters were brought to the surface ocean, their carbon isotope composition must not have reequilibrated with the atmosphere, given that old ventilation ages are evident during the YD, when the deep AM appears to have lost much of its glacial heat. Probably this was due to surface stratification, sea-ice cover, and insufficient time at the surface for equilibration [analogous to modern upwelling of Circumpolar Deep Water and Antarctic Bottom Water (AABW) formation]. Whereas previous studies have suggested that a subsurface incursion of warm Atlantic water during HS1 triggered the collapse of ice shelves (*39*), we speculate that the release of deep-ocean heat stored in the previously isolated deep AM may have also contributed to the melting of ice shelves and terminal ice-sheet margins (as well as sea ice and icebergs) in the circum-AM region during the end of the glacial period. The buildup and release of heat from an isolated glacial deep AM plausibly also played a role in earlier glacial-to-deglacial transitions.

REFERENCES AND NOTES

1. K. Aagaard, J. H. Swift, E. C. Carmack, *J. Geophys. Res.* **90**, 4833–4846 (1985).
2. C. Mauritzen, *Deep-Sea Res.* **43**, 769–806 (1996).
3. M. Sarnthein *et al.*, in *The Northern North Atlantic: A Changing Environment*, P. R. Schäfer, W. Ritzrau, M. Schlüter, J. Thiede, Eds. (Springer, Berlin, 2001), pp. 365–410.
4. S. Rahmstorf, *Nature* **419**, 207–214 (2002).
5. T. Dokken, E. Jansen, *Nature* **401**, 458–461 (1999).
6. S. Barker *et al.*, *Nature* **457**, 1097–1102 (2009).
7. J. D. Shakun *et al.*, *Nature* **484**, 49–54 (2012).
8. L. F. Robinson *et al.*, *Science* **310**, 1469–1473 (2005).
9. A. Burke, L. F. Robinson, *Science* **335**, 557–561 (2012).
10. L. C. Skinner, S. Fallon, C. Waelbroeck, E. Michel, S. Barker, *Science* **328**, 1147–1151 (2010).
11. L. C. Skinner, N. J. Shackleton, *Paleoceanography* **19**, PA2005 (2004).
12. D. J. R. Thornalley, S. Barker, W. S. Broecker, H. Elderfield, I. N. McCave, *Science* **331**, 202–205 (2011).
13. W. Broecker, T.-H. Peng, *Tracers in the Sea* (Lamont-Doherty Geological Observatory, Columbia Univ., Palisades, NY, 1982).
14. N. Thiagarajan, A. V. Subhas, J. R. Southon, J. M. Eiler, J. F. Adkins, *Nature* **511**, 75–78 (2014).
15. C. Cléroux, P. deMenocal, T. Guilderson, *Quat. Sci. Rev.* **30**, 1875–1882 (2011).
16. R. N. Sortor, D. C. Lund, *Earth Planet. Sci. Lett.* **310**, 65–72 (2011).
17. M. Sarnthein, P. Grootes, J. P. Kennett, M. J. Nadeau, in *Ocean Circulation: Mechanisms and Impacts*, A. Schmittner, J. Chang, S. Hemming, Eds. (AGU Geophysical Monograph Series vol. 173, American Geophysical Union, Washington, DC, 2007), pp. 175–196.
18. D. J. Wilson, K. C. Crocket, T. van de Flierdt, L. F. Robinson, J. F. Adkins, *Paleoceanography* **29**, 1072–1093 (2014).
19. H. A. Bauch *et al.*, *Quat. Sci. Rev.* **20**, 659–678 (2001).
20. L. Polyak *et al.*, *Global Planet. Change* **68**, 5–17 (2009).
21. J. Yu, H. Elderfield, A. M. Piotrowski, *Earth Planet. Sci. Lett.* **271**, 209–220 (2008).
22. M. Jakobsson, *Geochem. Geophys. Geosyst.* **3**, 1–18 (2002).
23. R. F. Spielhagen, *Polarforschung* **82**, 19–36 (2012).
24. D. Bauch, H. A. Bauch, *J. Geophys. Res. Oceans* **106**, 9135–9143 (2001).
25. R. Stein *et al.*, *Science* **264**, 692–696 (1994).
26. P. U. Clark *et al.*, *Science* **325**, 710–714 (2009).
27. T. P. Rippeth *et al.*, *Nat. Geosci.* **8**, 191–194 (2015).
28. R. L. Dickson, J. Lazier, J. Meinicke, P. Rhines, J. Swift, *Prog. Oceanogr.* **38**, 241–295 (1996).
29. W. B. Curry, D. Oppo, *Paleoceanography* **20**, PA1017 (2005).
30. E. P. Jones, B. Rudels, L. G. Anderson, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **42**, 737–760 (1995).
31. S. S. Hoffmann, J. F. McManus, W. B. Curry, L. S. Brown-Leger, *Nature* **497**, 603–606 (2013).
32. N. Nørgaard-Pedersen *et al.*, *Paleoceanography* **18**, 1063 (2003).
33. T. Veum, E. Jansen, M. Arnold, I. Beyer, J.-C. Duplessy, *Nature* **356**, 783–785 (1992).
34. U. Pflaumann *et al.*, *Paleoceanography* **18**, 1065 (2003).
35. T. M. Cronin *et al.*, *Nat. Geosci.* **5**, 631–634 (2012).
36. M. Y. Meland, T. M. Dokken, E. Jansen, K. Hevroy, *Paleoceanography* **23**, PA1210 (2008).
37. M. L. Timmermans, C. Garrett, E. Carmack, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **50**, 1305–1321 (2003).
38. J. F. Adkins, A. P. Ingersoll, C. Pasquero, *Quat. Sci. Rev.* **24**, 581–594 (2005).
39. S. A. Marcott *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13415–13419 (2011).
40. P. J. Reimer *et al.*, *Radiocarbon* **51**, 1111–1150 (2009).
41. T. L. Rasmussen, E. Thomsen, *Geophys. Res. Lett.* **36**, L01601 (2009).
42. C. H. Lear, E. M. Mawbey, Y. Rosenthal, *Paleoceanography* **25**, PA4215 (2010).
43. A. K. Tripati *et al.*, *Geochim. Cosmochim. Acta* **74**, 5697–5717 (2010).

ACKNOWLEDGMENTS

We thank L. Keigwin, T. Marchitto, J. McManus, and D. Oppo for discussions and comments on the manuscript. D.J.R.T. is grateful to T. Dokken for the suggestion to look to the AM for the source of the poorly ventilated water reconstructed from data collected south of Iceland. Funding was provided by a WHOI Ocean and Climate Change Institute (OCCI) scholarship and OCCI grant 27071264 (D.J.R.T.); WHOI OCCI and NSF grants OIA-1124880 and OCE-1357121 (G.G.); the WHOI J. Lamar Worzel Assistant Scientist Fund, the Penzance Endowed Fund in Support of Assistant Scientists, and NSF grant ANT-1246387 (W.G.); Marie Curie grant no. 298513 under the European Union's Seventh Framework Programme for Research (M.Z.); Australian Research Council Discovery Project grant DP140101393 (J.Y.); and Natural Environment Research Council grant NE/J008133/1 (S.B.). Research was planned and performed by D.J.R.T., except as follows: sample material, faunal counts, and *N. pachyderma* (s) δ^{18}O data from core MD992276 were provided by H.A.B.; W.G., M.Z., and S.M.B. conducted clumped isotope measurements; and G.G. conducted the idealized one-dimensional modeling. The manuscript was written by D.J.R.T., with comments and contributions from all authors. Data used in this study can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/706/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S5
References (*44*–*70*)

20 February 2015; accepted 8 July 2015
10.1126/science.aaa9554

QUANTUM OPTICS

Universal linear optics

Jacques Carolan,[1] Christopher Harrold,[1] Chris Sparrow,[1,2] Enrique Martín-López,[3] Nicholas J. Russell,[1] Joshua W. Silverstone,[1] Peter J. Shadbolt,[2] Nobuyuki Matsuda,[4] Manabu Oguma,[5] Mikitaka Itoh,[5] Graham D. Marshall,[1] Mark G. Thompson,[1] Jonathan C. F. Matthews,[1] Toshikazu Hashimoto,[5] Jeremy L. O'Brien,[1] Anthony Laing[1]*

Linear optics underpins fundamental tests of quantum mechanics and quantum technologies. We demonstrate a single reprogrammable optical circuit that is sufficient to implement all possible linear optical protocols up to the size of that circuit. Our six-mode universal system consists of a cascade of 15 Mach-Zehnder interferometers with 30 thermo-optic phase shifters integrated into a single photonic chip that is electrically and optically interfaced for arbitrary setting of all phase shifters, input of up to six photons, and their measurement with a 12-single-photon detector system. We programmed this system to implement heralded quantum logic and entangling gates, boson sampling with verification tests, and six-dimensional complex Hadamards. We implemented 100 Haar random unitaries with an average fidelity of 0.999 ± 0.001. Our system can be rapidly reprogrammed to implement these and any other linear optical protocol, pointing the way to applications across fundamental science and quantum technologies.

Photonics has been crucial in establishing the foundations of quantum mechanics (*1*) and more recently has pushed efforts in understanding new nonclassical computational possibilities. Typical protocols involve nonlinear operations, such as the generation of quantum states of light through optical frequency conversion (*2*, *3*), or measurement-induced nonlinearities for quantum logic gates (*4*), together with linear operations between optical modes so as to implement core processing functions (*5*). Encoding qubits in the polarization of photons has been particularly appealing for the ability to implement arbitrary linear operations on the two polarization modes by using a series of wave plates (*6*). For path encoding, the same operations can be mapped to a sequence of beamsplitters and phase shifters. Because any linear optical (LO) circuit is described by a unitary operator, and a specific array of basic two-mode operations is mathematically sufficient to implement any unitary operator on optical modes (*7*), it is theoretically possible to construct a single device with sufficient versatility to implement any possible LO operation up to the specified number of modes.

Here, we report the realization of a six-mode device that is completely reprogrammable and universal for LO. The versatility of this universal LO processor (LPU) is demonstrated for several quantum information protocols, including tasks that were previously not possible. We implemented heralded quantum logic gates at the heart of the circuit model of LO quantum computing (*4*) and new heralded entangling gates that underpin the measurement-based model of LO quantum computing (*8*–*10*), both of which are the first of their kind in integrated photonics. We performed 100 different boson sampling (*11*–*15*) experiments, with new verification protocols realized simultaneously. Last, we used multiparticle quantum interference to distinguish six-dimensional complex Hadamard operations, including newly discovered examples, in which full classification remains an open mathematical problem. The results presented required reconfiguration of this single device to implement ~1000 experiments.

Isolated quantum mechanical processes, including lossless LO circuits acting on photons, preserve orthogonality between input and output states and are therefore described by unitary operators. Although the relevant mathematics for parameterizing unitary matrices has been known for at least a century (*16*), the theoretical formulation of a LO circuit in which an arbitrary unitary operator could be realized is more recent (*7*). The full space of m-dimensional unitary matrices (or unitaries) can be parametrized as a product of $\approx m^2/2$ two-dimensional unitary primitives, each with two free parameters. Any given unitary then corresponds to a specific set of parameter values. An arbitrarily reconfigurable LO network would be realized by implementing these unitary primitives as two-mode Mach-Zehnder interferometers (MZIs) with two beamsplitters (or integrated directional couplers) and two phase shifters, so that any given LO network corresponds to a set of phase-shift values.

Such operation can be understood, with reference to the schematic shown in Fig. 1A, as a sequence of $m - 1$ subunitaries D_i (Fig. 1A, diagonal dashed boxes), each of which is guaranteed to enable transformation of the state of an input photon (Fig. 1A, left) in waveguide i to an arbitrary superposition of quantum mechanical amplitudes in all its output waveguides (Fig. 1A, right). The top MZI in each D_i is set to give the desired probability and phase for the input photon to output in waveguide i. The other parts of the quantum mechanical amplitude of the photon undergo similar operations in each successive MZI of a given D_i in order to set the amplitudes and phases in subsequent waveguides. Full reconfigurability is realized by feeding all output waveguides from one $D_{i'}$ into all but the uppermost input waveguide of the following $D_{i'-1}$ (*17*).

Realizing this scheme requires subwavelength stability and high-fidelity components to support both classical and quantum interference—possibilities opened up with integrated quantum photonics (*18*–*26*). Our LPU (Fig. 1B), made with planar lightwave circuit (PLC) technology (*27*, *28*), comprises an array of 30 silica-on-silicon waveguide directional couplers with 30 electronically controlled thermo-optic phase shifters to form a cascade of 15 MZIs across six modes (fabrication details are available in the supplementary materials, section S1). This LPU can implement any six-dimensional unitary operator up to undetected output phases and any three-dimensional nonunitary matrix (*17*) by using operator dilation (*29*). Here, we focus on several protocols at the forefront of quantum information science and technology.

Quantum gates

With the addition of single-photon sources and measurements, and rapid feed-forward of classical information, both the circuit (*4*) and measurement-based (*8*–*10*) models of digital quantum computing can be efficiently implemented with LO. Basic two-qubit processes are realized probabilistically with LO circuitry; therefore, a key requirement for scalability is that a successful operation is heralded by the detection of ancillary photons (ancillas) in order to signal that the processed photonic qubits are available for use in the larger architecture (*30*). These are typically referred to as "heralded gates."

We programmed our device to implement a new compact four-photon scheme suitable for measurement-based quantum computing, which generates the maximally entangled state of two photonic qubits upon the detection of another two ancilla photons (Fig. 2, A and B). For Bell states, measurements in common bases should be correlated, whereas measurements in different bases should be uncorrelated. We implemented both types of measurement (Fig. 2C), finding the mean statistical fidelity $\mathcal{F}_s = \sum_i \sqrt{p_i^{\text{exp}} \cdot p_i^{\text{th}}}$ (*31*) between all four experimental p^{exp} and theoretical p^{th} probability distributions to be $\mathcal{F}_s = 0.966 \pm 0.004$ (an explanation of error analysis is provided in the supplementary materials, section S2.1). These measurements were used to verify the entanglement of our state by calculating

[1]Centre for Quantum Photonics, H. H. Wills Physics Laboratory, and Department of Electrical and Electronic Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK. [2]Department of Physics, Imperial College London, London SW7 2AZ, UK. [3]Nokia Technologies, Broers Building, 21 J. J. Thomson Avenue, Cambridge CB3 0FA, UK. [4]NTT Basic Research Laboratories, Nippon Telegraph and Telephone (NTT) Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan. [5]NTT Device Technology Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan.
*Corresponding author. E-mail: anthony.laing@bristol.ac.uk

$E = 1/2(\langle \sigma_x \otimes \sigma_x \rangle + \langle \sigma_z \otimes \sigma_z \rangle)$, finding a value of $E = 0.673 \pm 0.031$, where $E > 1/2$ witnesses entanglement (a full description of gate implementation and analysis is provided in the supplementary materials, section S3.3) (*32*).

Next, the device was reprogrammed to realize a quantum logic gate designed to be scalable for the circuit model of quantum computing (Fig. 2, D and E). The gate receives four photons: Two photons are each encoded with a qubit, and two photons act as ancillas. Detecting one photon in each of the two heralding modes signals the implementation of a controlled-NOT (CNOT) operation (*4*, *33*, *34*) between the two photonic qubits: The logical state of the target qubit is flipped ($0 \leftrightarrow 1$) if the control qubit is in the state 1 and left unchanged if the control qubit is in state 0. We measured the logical truth table for this operation in the computational basis (Fig. 2F), finding its mean statistical fidelity averaged over all computational inputs to be $\mathcal{F}_s = 0.930 \pm 0.003$ with the ideal case (supplementary materials, section S3.4).

Although the truth table measures the map between logical basis states, capturing complete phase information requires full process tomography, which we were able to perform for an unheralded CNOT gate. With no ancillas, this unscalable gate requires two photons, which are consumed as part of its operation. Several examples of photonic chips specifically fabricated to implement such two-qubit gates (*21*, *35*, *36*) have been reported. To compare the performance of our universal processor against devices fabricated for a specific task, we implemented a two-photon unheralded CNOT gate (*37*–*39*) with single-qubit preparation and measurement capabilities (Fig. 2, G and H) and performed full quantum process tomography (Fig. 2I) (*40*). The process fidelity was found to be $\mathcal{F}_p = 0.909 \pm 0.001$, and the average gate fidelity was $\mathcal{F}_g = 0.927 \pm 0.001$, which is greater than those previously reported (gate implementation is provided in section S3.5 and quantum process tomography details are in section S2.4 of the supplementary materials) (*21*, *35*, *41*).

Combined with two-qubit operations, a small set of single-qubit gates (Fig. 2J), including the Hadamard ($\hat{H}$) and $\pi/8$ ($\hat{T}$) gates, are sufficient to realize a universal gate set for quantum computing (*42*). We implemented and performed full quantum process tomography for these two gates and for the three Pauli gates (Fig. 2K), finding an average process fidelity of $\mathcal{F}_p = 0.992 \pm 0.008$.

In multiphoton experiments, deviations from unit fidelity are primarily caused by imperfections in the photon source, such as reduced quantum interference between different pair creation events (*43*), and higher-order terms in the spontaneous parametric down-conversion (SPDC) process. To omit these effects and measure the performance of our LPU directly, we used one- and two-photon ensembles to recover the raw transfer matrix M (*44*) implemented by the device, making no assumptions about its unitarity. We calculated the circuit fidelity $\mathcal{F}_c = \mathrm{Tr}(|U^{\dagger} \cdot M|^2)/6$ with the intended unitary U. For the Bell state generator and the heralded and unheralded CNOT, we found $\mathcal{F}_c = 0.943 \pm 0.004$, 0.941 ± 0.018, and 0.939 ± 0.040, respectively (supplementary materials, section S2.3).

Implementing boson sampling

The realization of a large-scale quantum computer that demonstrates an intrinsic exponential advantage over classical machines would be in conflict with a foundational tenet in computer science: the extended Church Turing thesis (ECT). The ECT conjectures that all realistic physical systems can be efficiently simulated with a probabilistic Turing machine, or classical computer.

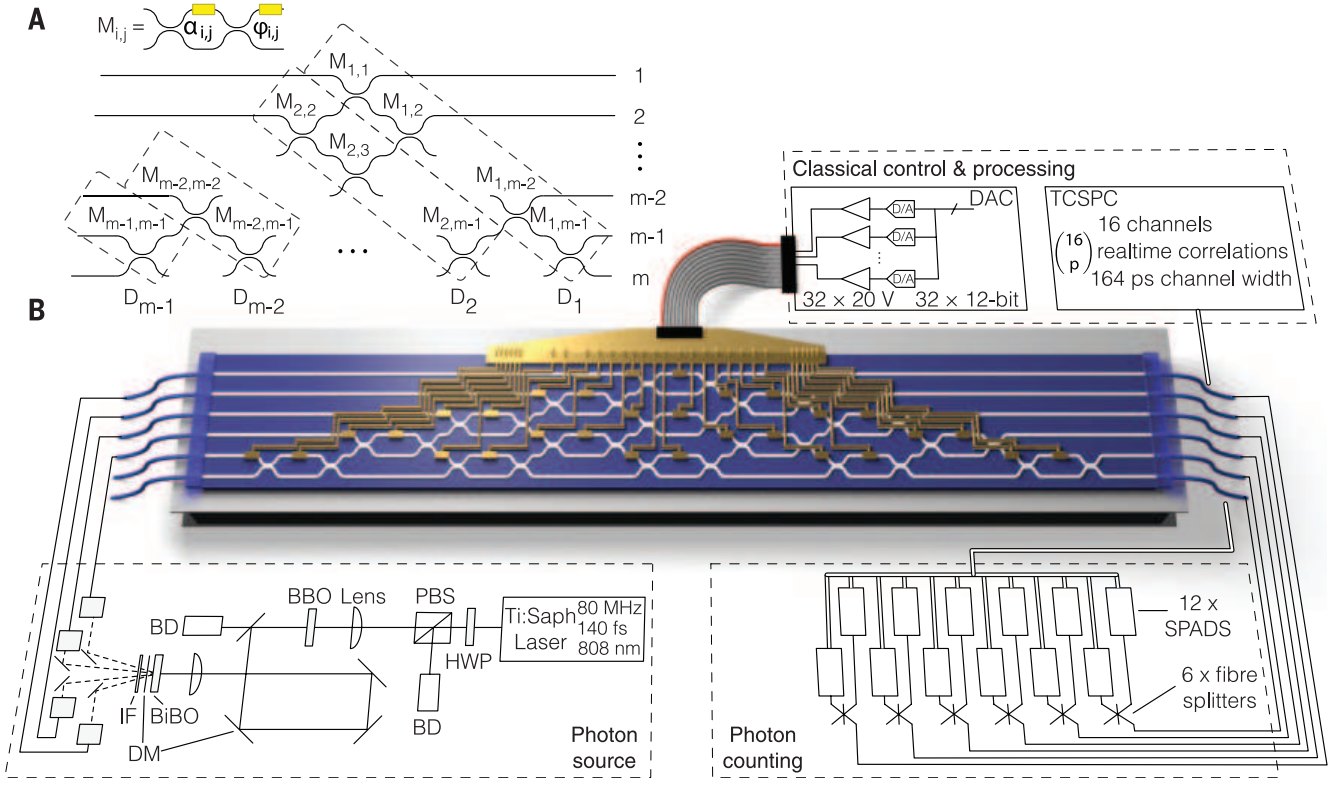


Fig. 1. Universal LPU. (**A**) Decomposition of a fully parametrized unitary for an m-mode circuit to realize any LO operation. Subunitaries D_i consist of MZIs $M_{i,j}$ built from phase shifters (yellow) and beam splitters, to control photon amplitudes ($\alpha_{i,j}$) and phases ($\varphi_{i,j}$). (**B**) Multiphoton ensembles are generated via SPDC, comprising a BiBO crystal, dichroic mirrors (DM), and interference filter (IF), preceded by a pulsed Ti:sapphire laser and second harmonic generation from a BBO crystal. Photons are collected into polarization-maintaining fibers and delivered to the LPU via a packaged v-groove fiber array (VGA). The processor is constructed over six modes as a cascade of 15 MZIs, controlled with 30 thermo-optic phase shifters, set with a digital-to-analog converter (DAC), and actively cooled with a Peltier cooling unit. Photons are then out-coupled into a second packaged VGA and sent to six (or 12 with fiber splitters for single-mode photon-number resolving capability) SPADs and counted by using a 12-channel time-correlated single-photon counting module (TCSPC) (supplementary materials, section S1).

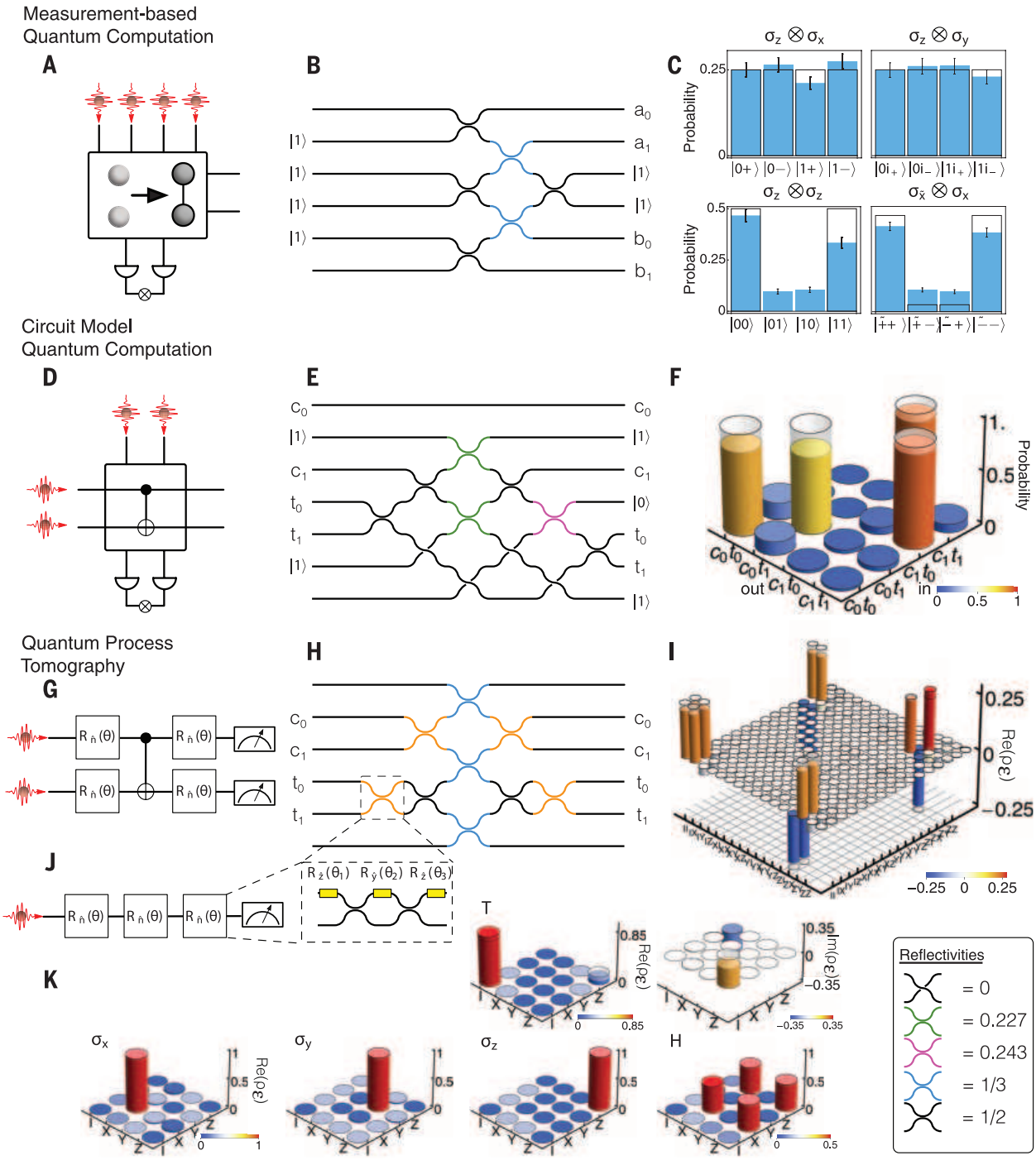


Fig. 2. Gates for LO quantum computing. (**A**) The heralded Bell state generator receives four photons and emits two of them in a maximally entangled state upon detection of the remaining two. (**B**) Our LO protocol emits a Bell state on modes {*a*, *b*}, with input and heralding modes labeled $|1\rangle$. (**C**) Experimental data measuring Bell state correlations in a given basis (blue), with ideal theoretical values overlaid. Error bars assume Poissonian counting statistics. (**D**) The heralded CNOT gate is successfully implemented on the two photonic qubits, upon detection of the two ancilla photons. (**E**) The LO protocol realizing the heralded CNOT operation on the control $c_{0,1}$ and target $t_{0,1}$ qubits. (**F**) Experimental data showing the computational truth table, with the ideal theoretical truth table overlaid. (**G**) Quantum process tomography of an unheralded two-qubit CNOT gate can be performed with the addition of arbitrary single-qubit preparation and measurement operations. (**H**) The LO circuit realizing the unheralded CNOT gate, with MZIs (inset) allowing single-qubit operations. (**I**) Experimentally determined process matrix, with ideal theoretical values overlaid. (**J**) An arbitrary single-qubit rotation $R_{\hat{n}}(\theta)$ about some vector ($\hat{n}$) on the Bloch sphere can be realized with a MZI and additional phase shifters. Three consecutive MZIs allow us to perform full-process tomography on any single-qubit operation. (**K**) Experimental data showing the measured process matrices for the three Pauli operations, the Hadamard gate ($\hat{H}$), and the $\pi/8$ phase gate ($\hat{T}$). Experimental data are corrected for measured detector efficiencies.

Designed for LO—with no requirement for quantum logic gates, qubit entangling operations, or number-resolved photon detection—boson sampling (*11*–*15*) is a quantum protocol that has been developed as a rapid route to challenge the ECT and demonstrate that quantum physics can be harnessed to provide fundamentally new and nonclassical computational capabilities.

Based on the foundations of computer science, boson sampling is a mathematical proof (using plausible conjectures) that a many-photon state, when acted on by a large LO circuit set to implement a Haar-random unitary, will give rise to a probability distribution that cannot be efficiently sampled by a classical algorithm. Quantum interference among the photons (*45*) contributes to the pattern of the probability distribution. The classical intractability arises because the probability amplitude for each correlated photon detection event is given by a classically hard function, known as the "permanent" (*46*), of the submatrix that describes a particular route of photons through the circuit. Experimentally, each detection event represents a sample drawn from that classically forbidden probability distribution.

Acting on three-photon ensembles, our device was programmed to implement 100 different boson sampling routines. Each circuit configuration was chosen randomly from the Haar measure, which was implemented via a direct parameterization of phase shifters (*47*), the probability

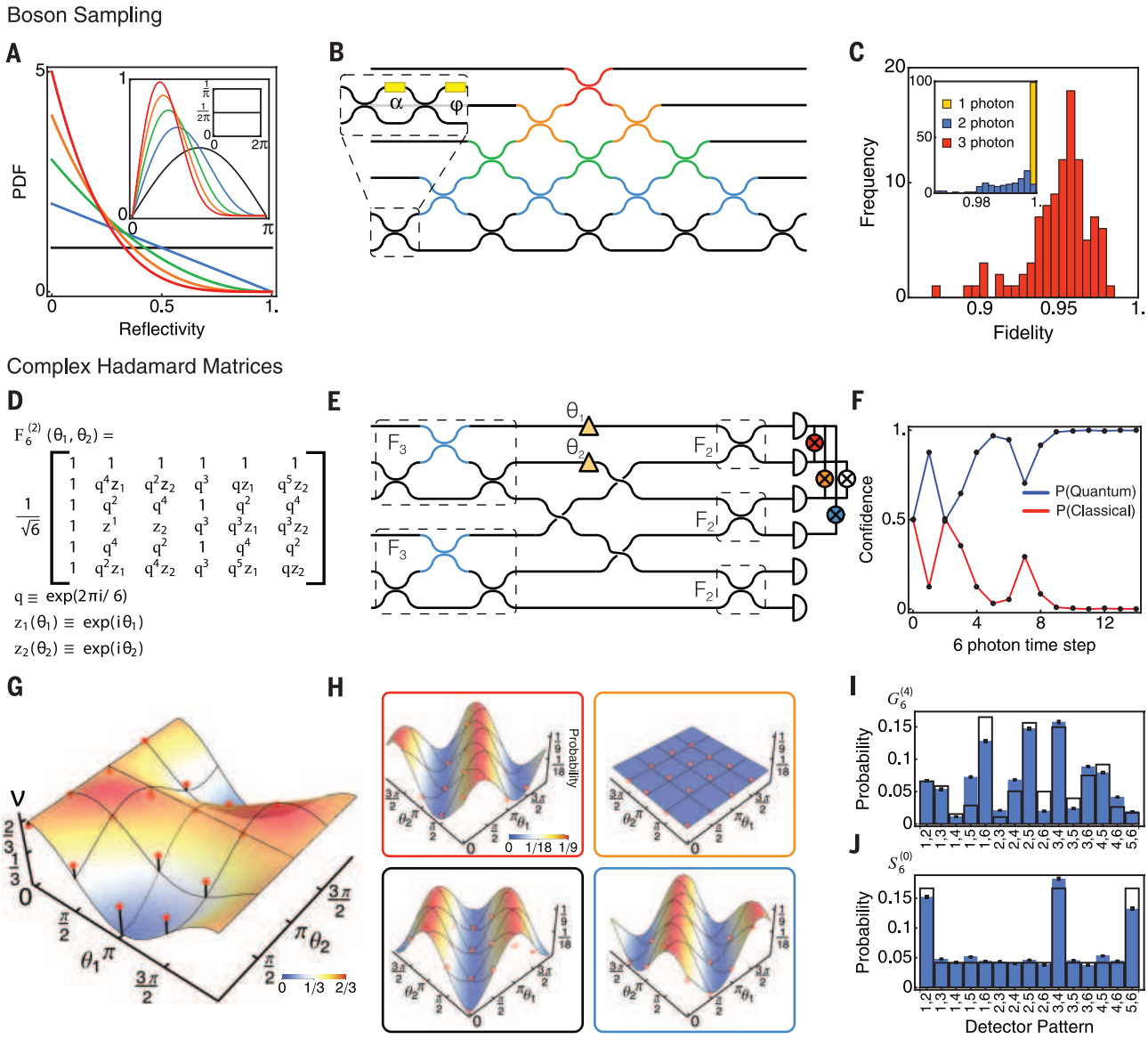


Fig. 3. Boson sampling and complex Hadamard matrices. (**A** and **B**) A Haar random unitary can be directly implemented by choosing beamsplitter reflectivities [or equivalently MZI phases α (inset 1)], and phase-shifter values φ (inset 2) from the probability density functions in (A). (**C**) A histogram of measured statistical fidelities for 100 three-photon boson sampling experiments, with one- and two-photon histograms inset ($\overline{F}_s$ = 0.999 ± 0.001, 0.990 ± 0.007, respectively). (**D** and **E**) The $F_6^{(2)}$ two-parameter (θ_1, θ_2) family of six-dimensional complex Hadamard matrices. (**F**) Dynamic updating of the confidence that six-photon detection events are sampled from a distribution of indistinguishable (quantum, blue) or distinguishable (classical, red) photons. (**G**) Three-photon violations (ν) of the zero-transmission law from scanning over $F_6^{(2)}(\theta_1, \theta_2)$. Experimental points in red are plotted with the ideal theoretical manifold. (**H**) Two-photon correlation manifolds in $F_6^{(2)}$ for the probability of a given detection event [as color coded in (E)], with experimental points in red. (**I** and **J**) The measured probability for a given detection event when two photons are injected into an instance of $G_6^{(4)}$ and $S_6^{(0)}$, with ideal theoretical black bars (error bars assume Poissonian counting statistics). Experimental data are corrected for measured detector efficiencies.

density functions of which are displayed in Fig. 3A. For each implementation (Fig. 3B), detection events were counted for each of the 20 collision-free ways in which three photons can exit the six output ports of the device. A histogram of fidelities is displayed in Fig. 3C with statistics based on calculations of matrix permanents, with a mean statistical fidelity of $\overline{\mathcal{F}_s}$ = 0.950 ± 0.020. These results demonstrate the performance of our LPU over many circuit configurations, randomly and unbiasedly chosen from the full space of all possible configurations.

Verifying boson sampling

An open and important question, particularly in light of the ECT, is how to verify that boson sampling continues to be governed by the laws of quantum mechanics when experiments reach the scale that classical computers can no longer simulate (*48*–*51*). Unlike certain algorithms for digital quantum computers—including Shor's factoring algorithm (*52*), whereby the solution to a problem believed to be classically hard can be efficiently checked—there seems to be no analogous way to check that large-scale boson sampling is sampling from a probability distribution that arises from many-photon quantum interference.

Although it is likely that boson sampling is in principle mathematically unverifiable, methods have been proposed to gather supporting or circumstantial evidence for the correct operation of the protocol (*51*, *53*, *54*). The essence of these methods is to implement experiments that share basic quantum mechanical features with boson sampling, but where certain properties of the experimental output can be predicted and therefore checked. The zero-transmission law (ZTL) predicts that correlated photon detection for most of the exponentially growing number of configurations is strictly suppressed if the circuit is set to implement the Fourier transform (FT) on optical modes (Fig. 3, D and E) (*55*). This is known because the structure of the FT allows these matrix permanents to be efficiently evaluated without explicit calculation. Because large-scale many-photon quantum interference is at the core of the ZTL, it has been proposed as a certificate for the capability of a device to implement boson sampling (*54*).

The LPU was programmed to implement 16 examples of the $F_6^{(2)}(\theta_1, \theta_2)$ two-parameter set of six-dimensional matrices, including F_6 (the six-dimensional FT), which occurs at $\theta_1,\theta_2 = \pi,0$. Using statistics from three-photon ensembles, the experimental violation of the ZTL was calculated as $\nu = N_s/N$, the ratio of the number of predicted suppressed events N_s to the total number of events N, the results of which are plotted in Fig. 3G alongside the theoretical manifold. The experimental points follow the shape of the manifold, with the minimal violation of the ZTL ν_{min} = 0.319 ± 0.009 occurring when F_6 is implemented. The average ZTL violation of the nine points that are predicted to maximally violate is $\overline{\nu}_{max}$ = 0.638 ± 0.029. Crucially, this verification protocol is implemented in the same device and with the same procedure as that for the boson sampling experiments above.

Six-photon verification

An essential requirement of boson sampling is that of indistinguishability among photons. With the LPU set to implement F_6, the six-photon state $|3_1, 3_2\rangle$ was injected, and six-photon statistics were counted with an all-fiber beam-splitter between each output mode and two single-photon avalanche diodes (SPADs) to give probabilistic number-resolved photon detection over a total of 12 SPADs. Although the complexity of states that are not one-photon-per-mode is less understood, the input state used here allows us to implement a protocol designed to verify indistinguishability among many photons with only a small number of detection events.

Bayesian model comparison was used to update, in real time, the confidence that events are sampled from a (precalculated) quantum probability distribution (arising from completely indistinguishable photons) or from a classical probability distribution (arising from completely distinguishable photons) (Fig. 3F) (*51*). After collecting 15 six-fold coincidence events, a confidence of P = 0.998 was determined that these are drawn from a quantum (not classical) distribution.

Complex Hadamard operations

The FT and $F_6^{(2)}$ are examples in the more general class of complex Hadamard matrices (CHMs), which are related to mutually unbiased bases (*56*) and are of fundamental interest in quantum information theory (*57*). CHMs are defined as $N \times N$ unitary matrices with entries of squared absolute value equal to $1/N$. Although this definition is straightforward, classification of these matrices is far from trivial and is concerned with identifying CHMs that are inequivalent up to pre- and postmultiplication with permutation matrices and diagonal unitaries (*58*). In the N = {2, 3, 5} case, all CHMs are equivalent to the respective FT matrix, whereas for N = 4, there exists a one-parameter equivalence class. Although a full classification of N = 6 CHMs is unknown, it is currently conjectured that the set consists of an isolated matrix $S_6^{(0)}$ that does not belong to any family (*59*) and a newly discovered four-parameter generic family $G_6^{(4)}$ (*60*).

In LO experimental implementations, discrimination among CHMs can be accomplished via the observation of characteristic patterns of photonic quantum interference (*61*–*64*). Up until now, these observations have been too experimentally challenging for the six-dimensional case. We reconstructed correlation manifolds of two-photon detection events by scanning over the $F_6^{(2)}$ matrices, displaying four (out of the 15 sets collected) in Fig. 3H. A mean statistical fidelity of $\overline{\mathcal{F}_s}$ = 0.979 ± 0.007 was found.

Last, we implemented an instance of $G_6^{(4)}$ [that is not contained in $F_6^{(2)}$] and $S_6^{(0)}$ and observed predicted characteristic two-photon quantum interference patterns (Fig. 3, I and J, respectively), with respective statistical fidelities of $\mathcal{F}_s$ = 0.986 ± 0.001 and $\mathcal{F}_s$ = 0.998 ± 0.001. The intractability of calculating the permanents of certain CHMs is an interesting research line, as is the possibility of searching for new CHMs by using photonic statistics.

Concluding remarks

Photonic approaches to quantum information science and technology promise new scientific discoveries and new applications. LO circuits lie at the heart of all of these protocols, and a single LPU device with the ability to arbitrarily "dial up" such operations promises to replace a multitude of existing and future prototype systems. Combining LPUs with existing higher-efficiency sources and detectors will expand their capabilities, and the development of LPUs with high-speed modulation (*65*) will enable the dynamically adaptive circuitry necessary for LO quantum computing. Integration of these components (*66*, *67*) with larger low-loss circuits (*68*) will open up new avenues of research and application.

REFERENCES AND NOTES

1. J.-W. Pan *et al.*, *Rev. Mod. Phys.* **84**, 777–838 (2012).
2. Z. Y. Ou, S. F. Pereira, H. J. Kimble, K. C. Peng, *Phys. Rev. Lett.* **68**, 3663–3666 (1992).
3. P. G. Kwiat *et al.*, *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
4. E. Knill, R. Laflamme, G. J. Milburn, *Nature* **409**, 46–52 (2001).
5. P. Kok *et al.*, *Rev. Mod. Phys.* **79**, 135–174 (2007).
6. N. K. Langford, thesis, University of Queensland (2007).
7. M. Reck, A. Zeilinger, H. J. Bernstein, P. Bertani, *Phys. Rev. Lett.* **73**, 58–61 (1994).
8. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
9. M. A. Nielsen, *Phys. Rev. Lett.* **93**, 040503 (2004).
10. D. E. Browne, T. Rudolph, *Phys. Rev. Lett.* **95**, 010501 (2005).
11. S. Aaronson, A. Arkhipov, in *Proceedings of the ACM Symposium on Theory of Computing* (ACM, New York, 2011), pp. 333–342.
12. M. A. Broome *et al.*, *Science* **339**, 794–798 (2013).
13. J. B. Spring *et al.*, *Science* **339**, 798–801 (2013).
14. M. Tillmann *et al.*, *Nat. Photonics* **7**, 540–544 (2013).
15. A. Crespi *et al.*, *Nat. Photonics* **7**, 545–549 (2013).
16. A. Hurwitz, *Nachr. Akad. Wiss. Gött. I. Math. Phys. Kl.* **1897**, 71 (1897).
17. D. A. B. Miller, *Photon. Res.* **1**, 1 (2013).
18. A. Politi, M. J. Cryan, J. G. Rarity, S. Yu, J. L. O'Brien, *Science* **320**, 646–649 (2008).
19. J. L. O'Brien, A. Furusawa, J. Vučković, *Nat. Photonics* **3**, 687–695 (2009).
20. G. D. Marshall *et al.*, *Opt. Exp.* **17**, 12546 (2009).
21. A. Crespi *et al.*, *Nat. Commun.* **2**, 566 (2011).
22. J. C. F. Matthews, A. Politi, A. Stefanov, J. L. O'Brien, *Nat. Photonics* **3**, 346–350 (2009).
23. B. J. Smith, D. Kundys, N. Thomas-Peter, P. G. R. Smith, I. A. Walmsley, *Opt. Exp.* **17**, 13516 (2009).
24. A. Laing *et al.*, *Appl. Phys. Lett.* **97**, 211109 (2010).
25. A. Peruzzo *et al.*, *Science* **329**, 1500–1503 (2010).
26. L. Sansoni *et al.*, *Phys. Rev. Lett.* **108**, 010502 (2012).
27. A. Himeno, K. Kato, T. Miya, *IEEE J. Sel. Top. Quantum Electron.* **4**, 913–924 (1998).
28. H. Takahashi, *Opt. Exp.* **19**, B173 (2011).
29. P. R. Halmos, *Summa Brasil. Math.* **2**, 125–134 (1950).
30. S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, A. Zeilinger, *Phys. Rev. Lett.* **93**, 020504 (2004).
31. M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2010).
32. H. Wunderlich, M. B. Plenio, *J. Mod. Opt.* **56**, 2100–2105 (2009).
33. T. C. Ralph, A. G. White, W. J. Munro, G. J. Milburn, *Phys. Rev. A* **65**, 012314 (2001).
34. R. Okamoto, J. L. O'Brien, H. F. Hofmann, S. Takeuchi, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10067–10071 (2011).
35. P. J. Shadbolt *et al.*, *Nat. Photonics* **6**, 45–49 (2011).
36. H. W. Li *et al.*, *New J. Phys.* **15**, 063017 (2013).

37. T. C. Ralph, N. K. Langford, T. B. Bell, A. G. White *Phys. Rev. A* **65**, 062324 (2002).
38. H. Hofmann, S. Takeuchi, *Phys. Rev. A* **66**, 024308 (2002).
39. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, D. Branning, *Nature* **426**, 264–267 (2003).
40. I. L. Chuang, M. A. Nielsen, *J. Mod. Opt.* **44**, 2455–2467 (1997).
41. J. L. O'Brien *et al*., *Phys. Rev. Lett.* **93**, 080502 (2004).
42. Y. Shi, *Quant. Inf. Comp.* **3**, 84 (2003).
43. M. Tanida, R. Okamoto, S. Takeuchi, *Opt. Exp.* **20**, 15275 (2012).
44. A. Laing, J. L. O'Brien, http://arxiv.org/pdf/1208.2868v1.pdf (2012).
45. C. K. Hong, Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
46. L. G. Valiant, *Theor. Comput. Sci.* **8**, 189–201 (1979).
47. N. J. Russell, J. L. O'Brien, A. Laing, http://arxiv.org/pdf/1506.06220.pdf (2015).
48. C. Gogolin, M. Kliesch, L. Aolita, J. Eisert, http://arxiv.org/pdf/1306.3995v2.pdf (2013).
49. S. Aaronson, A. Arkhipov, *Quant. Inf. Comp.* **14**, 1383 (2014).
50. N. Spagnolo *et al*., *Nat. Photonics* **8**, 615–620 (2014).
51. J. Carolan *et al*., *Nat. Photonics* **8**, 621–626 (2014).
52. P. W. Shor, *Proceedings of the 35th Annual Sympo-sium on Foundations of Computer Science* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994), pp. 124–134.
53. J. C. F. Matthews *et al*., *Sci. Rep.* **3**, 1539 (2013).
54. M. C. Tichy, K. Mayer, A. Buchleitner, K. Mølmer, *Phys. Rev. Lett.* **113**, 020502 (2014).
55. M. C. Tichy, M. Tiersch, F. de Melo, F. Mintert, A. Buchleitner, *Phys. Rev. Lett.* **104**, 220405 (2010).
56. I. Bengtsson *et al*., *J. Math. Phys.* **48**, 052106 (2007).
57. R. F. Werner, *J. Phys. Math. Gen.* **34**, 7081–7094 (2001).
58. W. Tadej, K. Życzkowski, *Open Syst. Inf. Dyn.* **13**, 133–177 (2006).
59. T. Tao, *Math Res. Lett.* **11**, 251–258 (2003).
60. F. Szollosi, *J. Lond. Math. Soc.* **85**, 616–632 (2012).
61. K. Mattle, M. Michler, H. Weinfurter, A. Zeilinger, M. Zukowski, *Appl. Phys. B* **60**, S111 (2004).
62. A. Peruzzo, A. Laing, A. Politi, T. Rudolph, J. L. O'Brien, *Nat. Commun.* **2**, 224 (2011).
63. A. Laing, T. Lawson, E. Martín López, J. L. O'Brien, *Phys. Rev. Lett.* **108**, 260505 (2012).
64. N. Spagnolo *et al*., *Nat. Commun.* **4**, 1606 (2013).
65. D. J. Thomson *et al*., *IEEE Photon. Technol. Lett.* **24**, 234–236 (2012).
66. J. W. Silverstone *et al*., *Nat. Photonics* **8**, 104–108 (2013).
67. J. P. Sprengers *et al*., *Appl. Phys. Lett.* **99**, 181110 (2011).
68. S. Sohma *et al*., *Optical Communications, 2006. ECOC 2006. European Conference on* (IEEE, New York, 2006), pp. 1–2.

ACKNOWLEDGMENTS

The authors acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC), the European Research Council (ERC), including BBOI, QUCHIP (H2020-FETPROACT-3-2014: Quantum simulation), PIQUE (FP7-PEOPLE-2013-ITN), the Centre for Nanoscience and Quantum Information (NSQI), the U.S. Army Research Office (ARO) grant W911NF-14-1-0133, and the U.S. Air Force Office of Scientific Research (AFOSR). J.C.F.M. was supported by a Leverhulme Trust Early Career Fellowship. J.L.O'B. acknowledges a Royal Society Wolfson Merit Award and a Royal Academy of Engineering Chair in Emerging Technologies.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/711/suppl/DC1
Materials and Methods
Tables S1 to S5
References (*69*–*72*)

16 April 2015; accepted 26 June 2015
Published online 9 July 2015
10.1126/science.aab3642

REPORTS

MAGNETISM

Imaging and control of ferromagnetism in $LaMnO_3/SrTiO_3$ heterostructures

X. Renshaw Wang,[1]*†‡ C. J. Li,[2,3]† W. M. Lü,[2] T. R. Paudel,[4] D. P. Leusink,[1] M. Hoek,[1] N. Poccia,[1] A. Vailionis,[5] T. Venkatesan,[2,3,6,7]* J. M. D. Coey,[2,8] E. Y. Tsymbal,[4] Ariando,[2,6] H. Hilgenkamp[1]

Oxide heterostructures often exhibit unusual physical properties that are absent in the constituent bulk materials. Here, we report an atomically sharp transition to a ferromagnetic phase when polar antiferromagnetic $LaMnO_3$ (001) films are grown on $SrTiO_3$ substrates. For a thickness of six unit cells or more, the $LaMnO_3$ film abruptly becomes ferromagnetic over its entire area, which is visualized by scanning superconducting quantum interference device microscopy. The transition is explained in terms of electronic reconstruction originating from the polar nature of the $LaMnO_3$ (001) films. Our results demonstrate that functionalities can be engineered in oxide films that are only a few atomic layers thick.

Modern thin-film deposition techniques enable the synthesis of complex oxide thin films with unit cell (uc)–level control over the thickness. Remarkably sharp phase transitions have been discovered in several systems upon increasing film thickness (*1*–*6*). The most prominent example is the two-dimensional electron gas formed between insulating thin films of $LaAlO_3$ and insulating TiO_2-terminated $SrTiO_3$ (STO) substrates, which occurs at a critical $LaAlO_3$ thickness of 4 uc (*2*). The possibility of selecting a different electronic or magnetic phase by adding a single layer of perovskite unit cells, with a lattice parameter of about 0.4 nm, offers tantalizing opportunities for nanostructured electronic and spintronic devices.

Because various interesting properties have been demonstrated in $LaMnO_3$ (LMO) bulk, thin films, and multilayers, ranging from orbital waves to its use as a catalyst for water splitting (*7*–*11*), LMO is an ideal candidate for observing control of functionalities in oxide heterostructures. It is a Mott insulator with an orthorhombic *Pbnm* structure, based on a $\sqrt{2}a_0$, $\sqrt{2}a_0$, $2a_0$ unit cell where $a_0 \approx 0.39$ nm is the elementary perovskite uc parameter. In stoichiometric LMO, La and Mn are both 3+ ions. The oxide is therefore a polar material that contains alternately charged $(LaO)^{1+}$ and $(MnO_2)^{1-}$ layers. Mn^{3+}, with electronic configuration $t_{2g}^3 e_g^1$ and spin $S = 2$, is a Jahn-Teller ion. If LMO had a perfect cubic perovskite structure, one would expect a conducting ground state, owing to the mobility of the unpaired electron in the degenerate e_g band. However, the e_g orbital degeneracy is lifted by the Jahn-Teller effect, and distorted MnO_6 octahedra line up with alternating long and short Mn-O bonds in the *a*-*b* plane, leading to orbital ordering, which results in electron localization (*12*). As a result of superexchange (*13*), the compound is an A-type antiferromagnet, with Mn spins ferromagnetically aligned within each plane and alternate (001) planes aligned antiferromagnetically. The Néel temperature is ~140 K. The antiferromagnetism is slightly canted in the bulk to produce a weak ferromagnetic moment of ~0.18 μ_B (where μ_B is the Bohr magneton) per uc (*14*–*16*) that is attributed to the antisymmetric Dzyaloshinskii-Moriya (DM) interaction associated with rotation of the MnO_6 octahedra. In thin films, ferromagnetism with a Curie temperature of ~115 K (*17*) accompanied by insulating behavior is often observed. The origin of this ferromagnetism is still unclear, but in addition to the above-mentioned DM mechanism, defects and epitaxial strain can be important factors (*17*–*20*). Here, we report a controllable monolayer-critical magnetic effect, whereby a uniform ferromagnetic state appears in $LaMnO_3$ at a critical thickness of 6 uc.

[1]MESA+ Institute for Nanotechnology, University of Twente, Enschede, Netherlands. [2]NUSNNI-Nanocore, National University of Singapore, Singapore. [3]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore. [4]Department of Physics and Astronomy and Nebraska Center for Materials and Nanoscience, University of Nebraska, Lincoln, NE, USA. [5]Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA, USA. [6]Department of Physics, National University of Singapore, Singapore. [7]Department of Electrical and Computer Engineering and Department of Materials Science and Engineering, National University of Singapore, Singapore. [8]School of Physics and Centre for Research on Adaptive Nanostructures and Nanodevices, Trinity College, Dublin, Ireland.
***Corresponding author. E-mail: wang.xiao@utwente.nl (X.R.W.); venky@nus.edu.sg (T.V.)** †These authors contributed equally to this work. ‡Present address: Massachusetts Institute of Technology, Cambridge, MA, USA.

In our study, LMO (001) films with thicknesses varying from 1 to 24 uc were grown by means of pulsed laser deposition monitored by reflection high-energy electron diffraction (RHEED), on TiO_2-terminated (001)-oriented STO substrates, in a 10^{-2}-mbar oxygen pressure. LMO films are coherently strained up to 20 uc, as verified by x-ray diffraction and RHEED. All the films were found to be insulating (*21*).

The distribution of local magnetic stray field emanating from the LMO films was imaged by scanning superconducting quantum interference device (SQUID) microscopy (SSM) (*22*) in zero applied field at 4.2 K. The essential part of the SSM is the square pickup loop with an inner size of ~$3 \times 5\ \mu m^2$ (Fig. 1A). During the measurement, the loop was scanned ~2 μm above the sample surface at a contact angle of ~10°. The SSM records the variation of magnetic flux threading the pickup loop, and the flux is converted to magnetic field by dividing by the effective pickup area of ~15 μm^2. The typical flux sensitivity of the SSM is around 14 $\mu\Phi_0\ Hz^{-1/2}$, where $\Phi_0 = 2 \times 10^{-15}\ Tm^2$ is the flux quantum and the bandwidth is 1000 Hz. Because our SSM sensor has a 10° inclination, the measured magnetic stray field component, denoted as B_z, is almost perpendicular to the sample surface. The SSM images are *x-y* maps of the magnetic-field values (which are converted to a color scale). The practical sensitivity during measurements set by external noise sources is estimated to be about 30 nT.

Figure 1B shows a typical scan of a 200 by 200 μm^2 area of a 6-uc LMO film grown on STO, with a pixel size of 1 by 1 μm^2. The cooling and measurement were both performed in zero applied field. An irregular pattern of regions with opposite magnetic field orientation was found, presenting a direct image of magnetic field emanating from ferromagnetic domains in the LMO. In Fig. 1, C and D, the local magnitude variations along two orthogonal directions within the plane are presented. Bulk magnetization measurements of ferromagnetic samples found a polarization of the film of 0.3 T (*21*), corresponding to an average moment of 1.6 μ_B per Mn atom. Because the spin-only moment on Mn^{3+} is 4 μ_B, the ferromagnetism appearing in films of LMO cannot be collinear.

Figure 2 shows that a monolayer-sharp transition to ferromagnetism occurs at a critical thickness of 6 uc. Figure 2, A to D, indicates the characteristic domain size evolution with thickness above the critical value. The SSM signals for films with a thickness smaller than the critical value (Fig. 2, E and F) are uniformly much weaker and cannot be resolved; they are two orders of magnitude smaller than the typical root-mean-square (RMS) values of the thicker films. The images for films below the critical thickness are shown on a color scale of one order of magnitude more sensitivity than that used for the ferromagnetic films. The critical thickness for ferromagnetism was confirmed by SSM measurements on another set of samples fabricated in a different growth chamber (*21*). Because a uniform and controllable ferromagnetic state is necessary for device application, the observed critical thickness for ferromagnetism is of particular importance. A critical thickness for ferromagnetism in Sr-doped LMO and $SrRuO_3$ thin films determined with bulk sensitive techniques has been reported (*3*, *4*). However, owing to the lack of domain structure information, characteristics and origin of the transition remain unclear. Using the SSM, Kalisky *et al.* (*6*) performed a comprehensive imaging study of magnetic structures on the interface between $LaAlO_3$ and STO below and above the critical thickness for ferromagnetism at the interface and observed magnetic patches above the critical thickness. Here, we observe an abrupt magnetic transition in which the whole film switches to a ferromagnetic state.

To find out whether the ferromagnetism is indeed solely dependent on the LMO thickness, we fabricated and measured a sample that consisted partly of a 5-uc-thick and partly of a 7-uc-thick LMO film. The sample was prepared by first growing the 5-uc LMO film and then covering half of the surface with a shadow mask while two extra uc of LMO were grown on the exposed surface. As seen from Fig. 3A, the 5-uc area exhibits no SSM signal within the noise level, and the 7-uc area reveals an inhomogeneous magnetic field distribution comparable to that of the regular 7-uc LMO films (Fig. 2C). Figure 3B compares RMS values of the SSM signal for films with different thicknesses; the difference in magnetic field value of films with thicknesses below and above the critical thickness is about two orders of magnitude.

The magnetization orientation of the ferromagnetic LMO films was determined by measuring the magnetization of a 7-uc LMO film grown on STO along different orientations with a vibrating sample magnetometer. During the measurement, the sample is first cooled in a 1-T magnetic field, then the moment is measured during warm-up in 0.1 T. Figure 3C shows the temperature-dependent magnetic moments of a 7-uc LMO film along two different orientations, revealing the in-plane nature of the magnetization. Therefore, the magnetic field pattern imaged by SSM is due to in-plane ferromagnetism. Magnetic moment as a function of temperature for LMO films with different thicknesses is shown in Fig. 3D; 4-uc and 5-uc LMO films show no clear sign of a magnetic moment, indicating an antiferromagnetic or a nonmagnetic state. Because these thinner films show no sign of either ferromagnetism or Curie-law paramagnetism (fig. S10F) in bulk magnetic measurements, yet strong magnetic moments on the Mn sites are still expected to be present, it is likely that the thin LMO films are antiferromagnetic, similar to the bulk (*21*). Furthermore, the

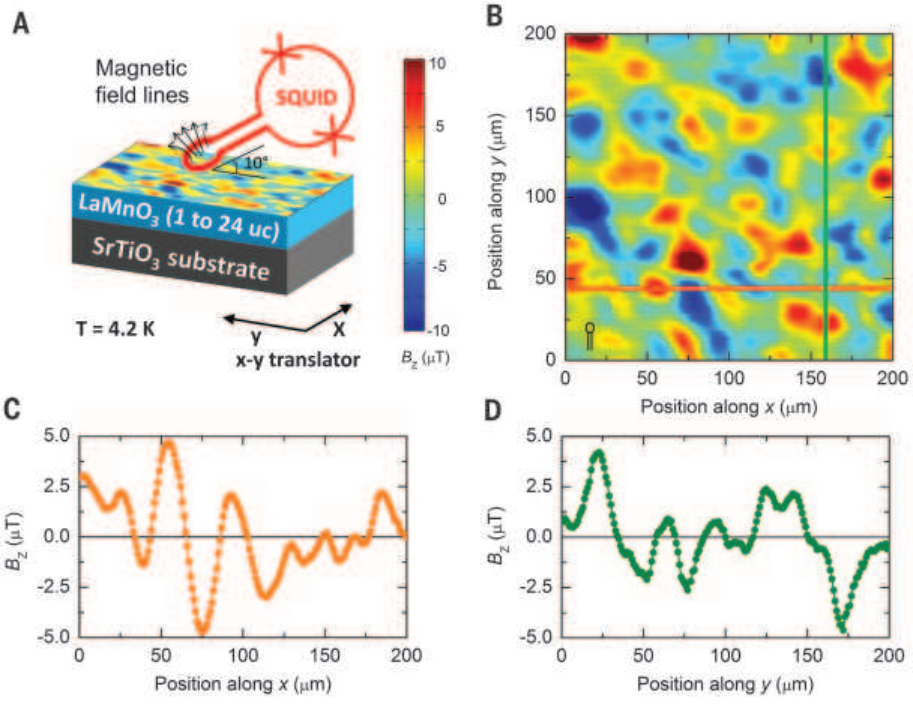


Fig. 1. Scanning SQUID microscopy on a 6-uc LMO film on a STO substrate. (**A**) Schematic of the microscopy technique with sketch of the pickup loop (red). (**B**) Image of the inhomogeneous stray field distribution of a 6-uc LMO film at 4.2 K. The red-yellow peaks in the two-dimensional color map indicate regions where the magnetic stray field exits the sample, and the blue peaks indicate regions where the magnetic stray field enters into the sample. The scan direction is always horizontally from left to right. A sketch (black) in the bottom-left corner of (B) indicates the size of the pickup loop. (**C** and **D**) *x*- (C) and *y*-direction (D) magnetic profiles along the orange and green lines in (B).

absence of an uncompensated magnetic moment in LMO films with an odd number of unit cells less than 6 (for example, 5) suggests antiferromagnetic order within (001) planes of the sample, rather than the planar antiferromagnetism that is found in the bulk (*21*).

Because of the abruptness of the phase transition and the similarity in polar properties of LMO (001) and $LaAlO_3$ (001) films, it is relevant to consider possible electronic reconstruction, which has been proposed as the mechanism for the abrupt insulator-to-metal transition in the well-studied $SrTiO_3$-$LaAlO_3$ case (*23*). LMO contains alternately charged $(LaO)^{1+}$ and $(MnO_2)^{1-}$ layers, resulting in an internal electric field E_0. A simple first-order estimate for this field is $E_0 = e/2A\varepsilon_0\varepsilon_r$, where e is the elementary charge, A is the unit cell area, ε_0 is the permittivity of vacuum, and ε_r is the dielectric constant of LMO. Taking $\varepsilon_r \sim 70$ at low temperature (*24*), the value of E_0 is 0.85 eV/nm, which will shear both the valence and conduction bands (Fig. 4A). The band gap (E_g) in bulk LMO [about 1.3 eV (*25*)] is smaller than that in bulk STO (3.2 eV), and therefore, charge transfer to eliminate the polar discontinuity occurs entirely within the polar LMO film. This is quite unlike the situation for LAO/STO, where the larger band gap of LAO (5.6 eV) ensures that the electrons are transferred to the STO side of the interface. At a certain thickness $t_c = E_g/E_0 \approx 4$ uc (1.53 nm), the valence band maximum of the LMO reaches the conduction band minimum at the LMO/STO interface, initiating electron transfer from the top to the bottom of the LMO film. Such transfer then decreases the electric field in LMO. The transferred charge as a function of thickness is zero below t_c and increases asymptotically with increasing thickness to 0.5*e* (*21*, *26*). The electron transfer therefore makes the interface region of LMO electron-doped and the top surface region of LMO hole-doped. This is analogous to the electronic phase separation in bulk manganites (*27*), which in our case is stabilized by an intrinsic electric field of the polar LMO film. The doping of LMO tilts the exchange interaction between Mn ions from superexchange to double exchange and leads to ferromagnetism when a sufficient number of electrons are available. Our density-functional calculations predict that both electron and hole doping of LMO increases the relative energy difference between the antiferromagnetic and ferromagnetic ground states (fig. S8), in agreement with the experimentally reported bulk phase diagram of LMO (*27*–*29*). Thus, electronic reconstruction appears to favor ferromagnetism above a certain critical thickness of LMO. We note that, contrary to the LAO/STO system, where the interface becomes metallic above the critical thickness, our films remain insulating. This is due to the difference in band gap between LAO and LMO. The large band gap of LAO leads to electron transfer to STO, whereas in the LMO/STO case, the electronic reconstruction results in the self-doping of LMO. The self-doping picture is also consistent with a previously reported unaltered Ti^{4+} valence state in $La_{1-x}Sr_xMnO_3$/STO heterojunctions (*29*).

The exact thickness at which the ferromagnetism occurs depends on the spread of the charge. Figure 4B shows an estimated doping charge per Mn atom when all the charge is projected onto just one uc (black curve) and when the charge is spread over 2 uc (red curve). We see that in the latter case, the bulk phase diagram predicts a phase transition from the insulating antiferromagnetic phase to the insulating ferromagnetic phase when the LMO thickness exceeds 6 uc, which is consistent with our experimental observations. The assumption of the charge spread over 2 uc agrees with results of first-principles calculations (*30*) and experimental observations (*29*) that indicate screening of the interface charge in doped LMO.

To verify the feasibility of this electronic reconstruction mechanism, we compared the behavior of 12-uc LMO and $CaMnO_3$ films, because the $Ca^{2+}Mn^{4+}O_3$ (001) film is nonpolar. These films were grown on TiO_2-terminated conducting (001)-oriented 0.1 weight % Nb-doped STO (Nb:STO) substrates and were covered by 2 uc $LaAlO_3$ under the same growth conditions. The 2-uc $LaAlO_3$ capping layer was used to reveal the effect of surface symmetry breaking, and the conducting Nb:STO substrate was used to investigate interface band-bending. By comparing SSM images between $LaAlO_3$-capped LMO grown on Nb:STO (Fig. 4C) and LMO grown on STO (Fig. 2B), we conclude that neither the surface symmetry breaking nor band-bending effects contribute to the critical magnetic behavior. Furthermore, as shown in Fig. 4, C and D, only the LMO films show the magnetic field patterns, whereas the nonpolar $CaMnO_3$ films show no signature of ferromagnetism. Thus, both our theoretical analysis and experimental data suggest

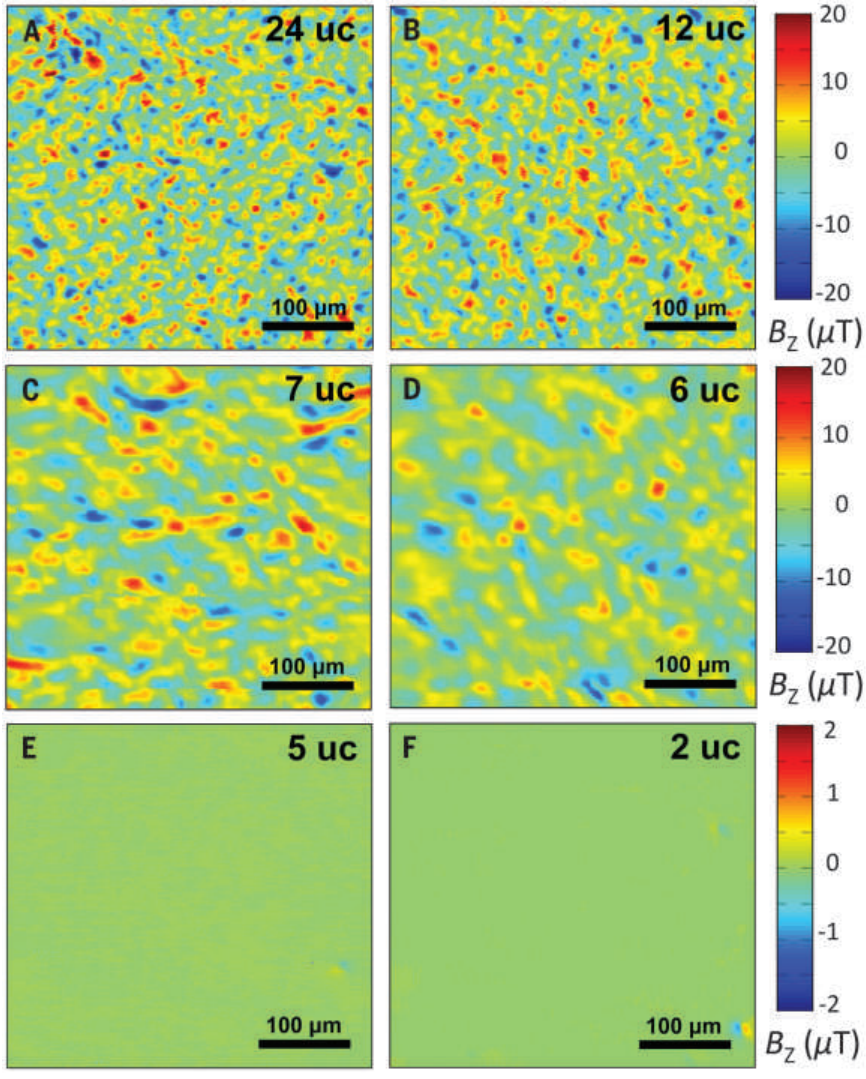


Fig. 2. Critical thickness for ferromagnetism in insulating LMO (001) films grown on STO substrates. All measurements were taken at 4.2 K. (**A** to **D**) Images of magnetic field emanating from LMO films with a thickness of (A) 24 uc, (B) 12 uc, (C) 7 uc, and (D) 6 uc, respectively. (**E** and **F**) Absence of magnetic field for 5-uc (E) and 2 uc (F) LMO. The scale of color bars for 5-uc and 2-uc LMO is one order of magnitude smaller than those of the other images.

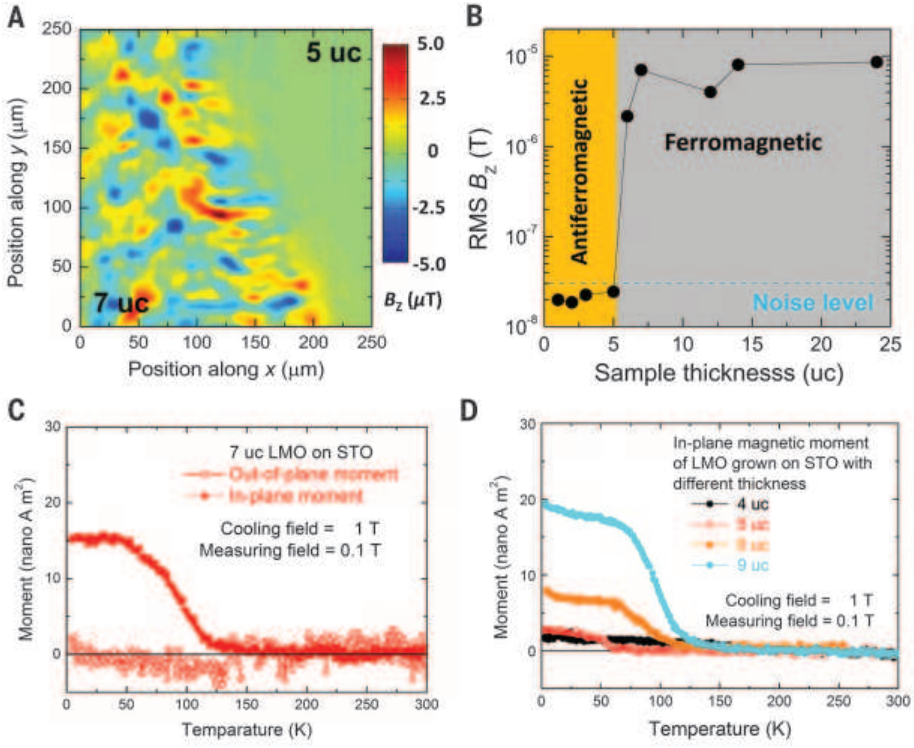


Fig. 3. Abrupt appearance of in-plane ferromagnetism. (**A**) Field distribution measured at 4.2 K in a sample with an upper-right area of 5-uc-thick LMO and a bottom-left area covered by 7-uc LMO. (**B**) RMS values of magnetic field for films with different thicknesses. (**C**) In-plane and out-of-plane magnetic moments of 7-uc LMO grown on STO. The magnetic moment of 7 uc of LMO is found to lie in-plane. (**D**) Magnetic moment of 4-, 5-, 6-, and 9-uc LMO films grown on STO as a function of temperature.

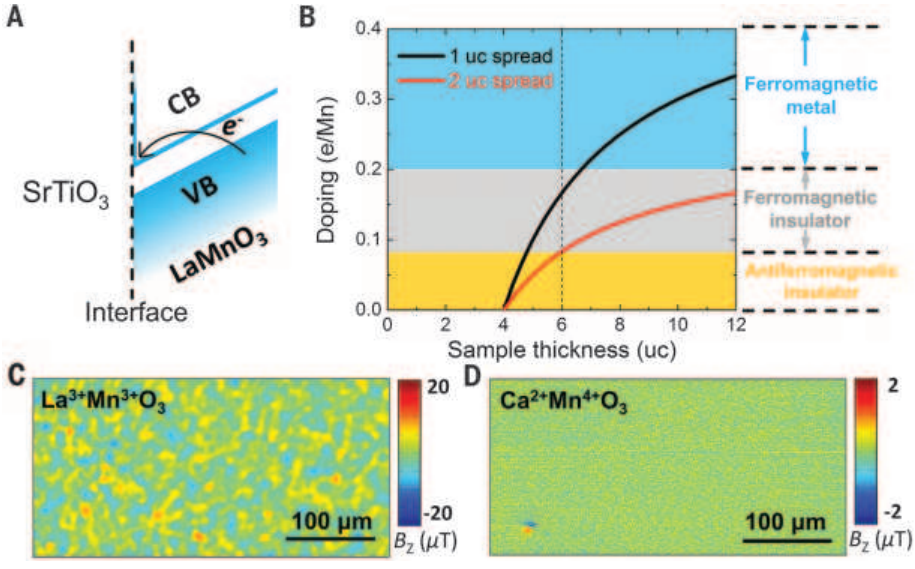


Fig. 4. Analysis of the origin of ferromagnetism in LMO. (**A**) Sketch of the band diagram of electronic reconstruction for LMO (001) film grown on an STO substrate (CB, conduction band; VB, valence band). (**B**) Amount of charge on the Mn site transferred from surface to interfacial layers as a function of LMO thickness. The black curve corresponds to the doping level where all the charge is projected onto just one uc, and the red curve corresponds to the doping level where the charge is spread over 2 uc. The three doping regimes corresponding to the ferromagnetic conducting state (blue area), the ferromagnetic insulating state (gray area), and the antiferromagnetic insulating state (orange area) are adapted from literature for doped bulk LMO (*27*, *28*). (**C**) SSM image of 12-uc $La^{3+}Mn^{3+}O_3$ showing ferromagnetic behavior. (**D**) SSM image of 12-uc $Ca^{2+}Mn^{4+}O_3$ shows no sign of ferromagnetic behavior.

that electronic reconstruction driven by the polar nature of LMO (001) films is a feasible explanation for the abrupt transition to ferromagnetic order in LMO thin films.

REFERENCES AND NOTES

1. J. Junquera, P. Ghosez, *Nature* **422**, 506–509 (2003).
2. S. Thiel, G. Hammerl, A. Schmehl, C. W. Schneider, J. Mannhart, *Science* **313**, 1942–1945 (2006).
3. M. Huijben *et al.*, *Phys. Rev. B* **78**, 094413 (2008).
4. J. Xia, W. Siemons, G. Koster, M. R. Beasley, A. Kapitulnik, *Phys. Rev. B* **79**, 140407 (2009).
5. A. V. Boris *et al.*, *Science* **332**, 937–940 (2011).
6. B. Kalisky *et al.*, *Nat. Commun.* **3**, 922 (2012).
7. E. Saitoh *et al.*, *Nature* **410**, 180–183 (2001).
8. C. H. Kim, G. Qi, K. Dahlberg, W. Li, *Science* **327**, 1624–1627 (2010).
9. J. Suntivich *et al.*, *Nat. Chem.* **3**, 546–550 (2011).
10. M. Gibert, P. Zubko, R. Scherwitzl, J. Iñiguez, J. M. Triscone, *Nat. Mater.* **11**, 195–198 (2012).
11. E. J. Monkman *et al.*, *Nat. Mater.* **11**, 855–859 (2012).
12. H. Sawada, Y. Morikawa, K. Terakura, N. Hamada, *Phys. Rev. B* **56**, 12154–12160 (1997).
13. J. B. Goodenough, *Phys. Rev.* **100**, 564–573 (1955).
14. G. Matsumoto, *J. Phys. Soc. Jpn.* **29**, 606–615 (1970).
15. J. M. D. Coey, M. Viret, S. von Molnar, *Adv. Phys.* **48**, 167–293 (1999).
16. V. Skumryev *et al.*, *Eur. Phys. J. B* **11**, 401–406 (1999).
17. A. Gupta *et al.*, *Appl. Phys. Lett.* **67**, 3494 (1995).
18. A. B. Shah *et al.*, *Adv. Mater.* **22**, 1156–1160 (2010).
19. J. Garcia-Barriocanal *et al.*, *Adv. Mater.* **22**, 627–632 (2010).
20. M. A. Roldan *et al.*, *Microsc. Microanal.* **19**, 1916–1917 (2013).
21. Information on materials and methods is available on *Science* Online.
22. J. R. Kirtley, J. P. Wikswo Jr., *Annu. Rev. Mater. Sci.* **29**, 117–148 (1999).
23. N. Nakagawa, H. Y. Hwang, D. A. Muller, *Nat. Mater.* **5**, 204–209 (2006).
24. P. Mondal, D. Bhattacharya, P. Choudhury, P. Mandal, *Phys. Rev. B* **76**, 172403 (2007).
25. A. Chainani, M. Mathew, D. D. Sarma, *Phys. Rev. B Condens. Matter* **47**, 15397–15403 (1993).
26. W. Son, E. Cho, B. Lee, J. Lee, S. Han, *Phys. Rev. B* **79**, 245411 (2009).
27. E. Dagotto, T. Hotta, A. Moreo, *Phys. Rep.* **344**, 1–153 (2001).
28. H. Fujishiro, T. Fukase, M. Ikebe, *J. Phys. Soc. Jpn.* **67**, 2582–2585 (1998).
29. J. A. Mundy *et al.*, *Nat. Commun.* **5**, 3464 (2014).
30. J. D. Burton, E. Y. Tsymbal, *Phys. Rev. Lett.* **106**, 157203 (2011).

ACKNOWLEDGMENTS

We thank M. S. Golden, A. Brinkman, D. Schlom, M. Huijben, Y. L. Zhao, and W. G. van der Wiel for valuable discussions; D. Veldhuis for assistance in SQUID design and fabrication; and J. R. Kirtley and the IBM T. J. Watson Research Center for providing the microscopy setup. This research was financially supported by the Dutch Foundation for Fundamental Research on Matter (FOM) foundation and through Netherlands Organisation for Scientific Research (NWO)–VICI and Rubicon grants (2011, 680-50-1114). The work at NUS is supported by the Singapore National Research Foundation (NRF) under the Competitive Research Programs (CRP Award no. NRF-CRP 8-2011-06 and CRP Award no. NRF-CRP10-2012-02) and the NUS FRC (AcRF Tier 1 grant no. R-144-000-346-112). Parts of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The work at University of Nebraska was supported by the NSF through Materials Research Science and Engineering Center (MRSEC, grant no. DMR-1420645). J.M.D.C. acknowledges support from Science Foundation Ireland grant 10/IN1.13006 and the European Union FP7 IFOX Project. All data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/716/suppl/DC1
Materials and Methods
Figs. S1 to S10
References (*31*–*50*)

17 December 2014; accepted 13 July 2015
10.1126/science.aaa5198

GEOPHYSICS

Layered deformation in the Taiwan orogen

T.-Y. Huang,[1] Y. Gung,[1*] B.-Y. Kuo,[2*] L.-Y. Chiao,[3] Y.-N. Chen[3]

The underthrusting of continental crust during mountain building is an issue of debate for orogens at convergent continental margins. We report three-dimensional seismic anisotropic tomography of Taiwan that shows a nearly 90° rotation of anisotropic fabrics across a 10- to 20-kilometer depth, consistent with the presence of two layers of deformation. The upper crust is dominated by collision-related compressional deformation, whereas the lower crust of Taiwan, mostly the crust of the subducted Eurasian plate, is dominated by convergence-parallel shear deformation. We interpret this lower crustal shearing as driven by the continuous sinking of the Eurasian mantle lithosphere when the surface of the subducted plate is coupled with the orogen. The two-layer deformation clearly defines the role of subduction in the formation of the Taiwan mountain belt.

The subduction of plate not only brings about continental or arc-continent collision but also affects mountain building in a complex way during the collision (*1*–*3*). The orogeny of Taiwan was preceded by the subduction of the Eurasian Plate (EP) eastward under the Philippine Sea plate at the Manila trench. With the consumption of the oceanic lithosphere in between, the Luzon arc approached the Eurasian continent at 4 to 5 million years ago (Ma) and initiated the collision (*4*) (Fig. 1). How the subduction of the EP interacts with the collision is a subject of debate. The thin-skinned critical taper wedge model asserts that the mountain is shaped by deformation above a detachment surface (décollement), below which the material is subducting along with the EP and intact from the collision (*5*, *6*). A thick-skinned lithospheric collision model advocates vertically continuous deformation on the lithospheric scale (*7*–*9*), thus including the subducted plate. Shear-wave splitting observations with large delay times in Taiwan, as in many other mountain belts, requires a coherent deformation, even extending to the sublithospheric mantle (*10*–*12*). This interpretation is open to question because shear-wave splitting measurements lack the adequate vertical resolution. Recently, the large-scale experiment Taiwan Integrated Geodynamics Research (TAIGER) (*8*, *9*) in the Taiwan collision zone has yielded improved observations of deformational structures in the lower crust, reinforcing the view of thick-skinned tectonics. However, this work established that deformation is not restricted to the upper crust but fails to define the deformation fabrics in details that could help to infer the interaction between the collision and subduction.

We developed three-dimensional (3D) tomographic models of shear-wave velocity (*Vs*) and its azimuthal variation for the crust of Taiwan, using the empirical Green's functions (EGF) of

[1]Department of Geosciences, National Taiwan University, Taipei, Taiwan. [2]Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan. [3]Institute of Oceanography, National Taiwan University, Taipei, Taiwan.
*Corresponding author. E-mail: ycgung@ntu.edu.tw (Y.G.); byk@earth.sinica.edu.tw (B.-Y.K.)

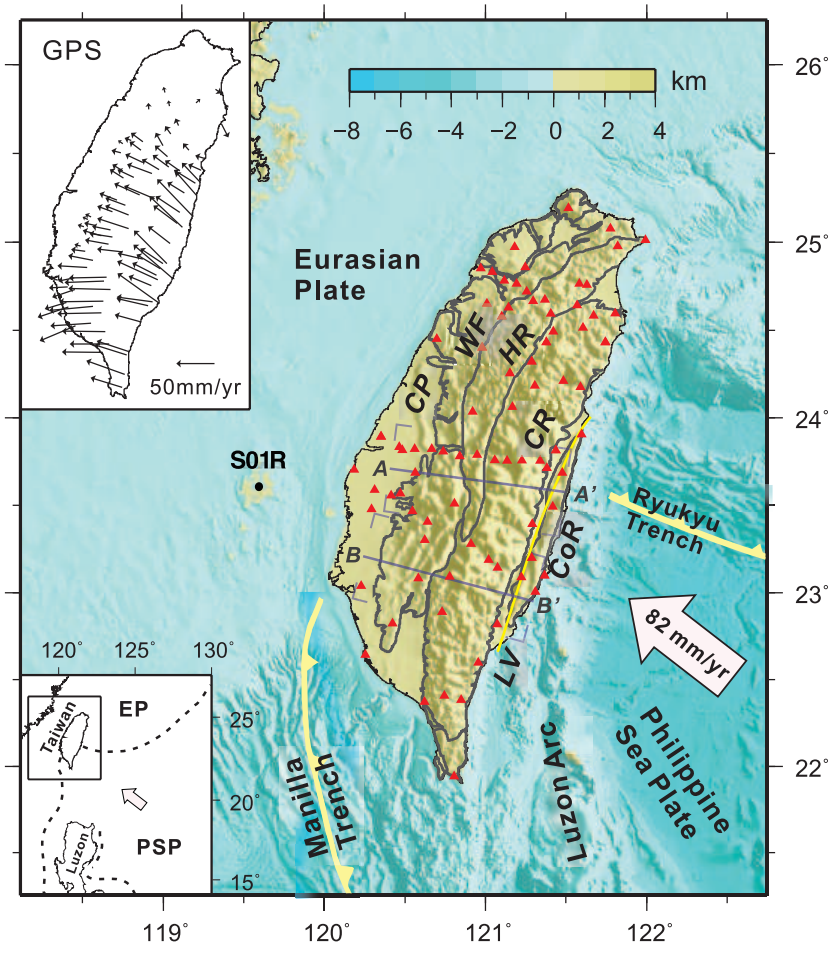


Fig. 1. The Taiwan orogen and regional plate tectonics. The convergence of the Philippine Sea plate (PSP) toward the EP at 82 mm/year (arrow) is being absorbed by the orogeny at Taiwan and the subduction at the Ryukyu trench. The GPS field (*27*) (simplified, upper left inset) illustrates the motion of the Taiwan orogen relative to the reference station S01R at the EP. Two cross sections, AA′ and BB′, sample the mature and young portions of the mountain belt (average area bracketed), respectively. Broadband seismic stations used in the ambient noise tomography are denoted by red triangles. LV, Longitudinal Valley or the suture zone; CoR, Coastal Range (*26*).

Rayleigh waves extracted from ambient seismic noises. With the unprecedented data amount and various advantages of noise-derived EGFs over conventional earthquake-based surface waves (*13*), we are able to resolve the detailed variations of seismic anisotropy from the shallow crust to about a 30-km depth. We implemented a wavelet-based multiscale inversion technique (*14*) to solve the simultaneous inversion for both isotropic and anisotropic models (*15*).

The tomographic models demonstrate that both *V*s and azimuthal anisotropy at shallow depths correlate well with the surface geology (Fig. 2). Among geological provinces of the Taiwan orogen (Fig. 1), the Coastal Plain (CP) and the Western Foothills (WF) are characterized by low *V*s, whereas the Hsuehshan Range (HR) and the Central Range (CR) are underlain by high *V*s. The fast directions of azimuthal anisotropy are generally parallel to the trend of the mountain belt, from north-to-south in southern Taiwan to northeast-to-southwest in northern Taiwan. This orogen-parallel anisotropy (OPA) may reflect the structural fabric of the brittle crust under compression, which can be the foliation of the metamorphic rocks of the HR and CR and the north-south aligned fold and thrust structures in the WF. These shallow crustal features, when combined with the teleseismic shear-wave splitting observations (*11*), support the hypothesis that orogenic processes generate the same deformational fabrics down to the sublithospheric mantle (*9*, *10*).

The close correlation of seismic properties with surface geology and structural trends diminishes with depth. At depths greater than 10 to 20 km, lower *V*s is associated with the HR and CR, and the fast axis of anisotropy aligns with the global positioning system (GPS)–indicated convergence direction. This convergence-parallel anisotropy (CPA) increases to ~30-km depth, below which the model resolution degrades. The same pattern of anisotropy may persist deeper as the azimuthal variation in the velocity of P_n, the *P* wave propagating in the uppermost mantle, exhibits similar east-west fast directions (*16*).

Two example cross sections averaging along central and southern Taiwan show variations in the fast direction of anisotropy with depth (Fig. 3). We model the high-angle rotation of anisotropy as a function of depth, using a scaled arctangent function, and define the anisotropy transition boundary (ATB) between the OPA and the CPA and the top of the CPA (fig. S4) (*15*). The ATB tilts gently from 5- to 10-km depths beneath the CP and WF to a 15-km depth beneath the CR. The CPA develops 2 to 5 km below the ATB in the upper crust in western Taiwan and in the lower crust beneath the CR. We performed resolution and recovery tests to show the robustness of a sharp ATB (Fig. 3) (*15*).

We propose that the CPA results from alignment of anisotropic minerals by subhorizontal shearing parallel to the convergence direction and that the ATB separates the compression-dominant upper crust and the shear-dominant mid- to lower crust. Amphibole, an abundant mineral in the lower crust (*17*), may be the candidate anisotropic agent potentially capable of producing high azimuthal anisotropy with fast direction parallel to shear (*18*). The lineation on subhorizontal rock cleavages aligned by shearing may also give rise to CPA (*19*). The depth distribution

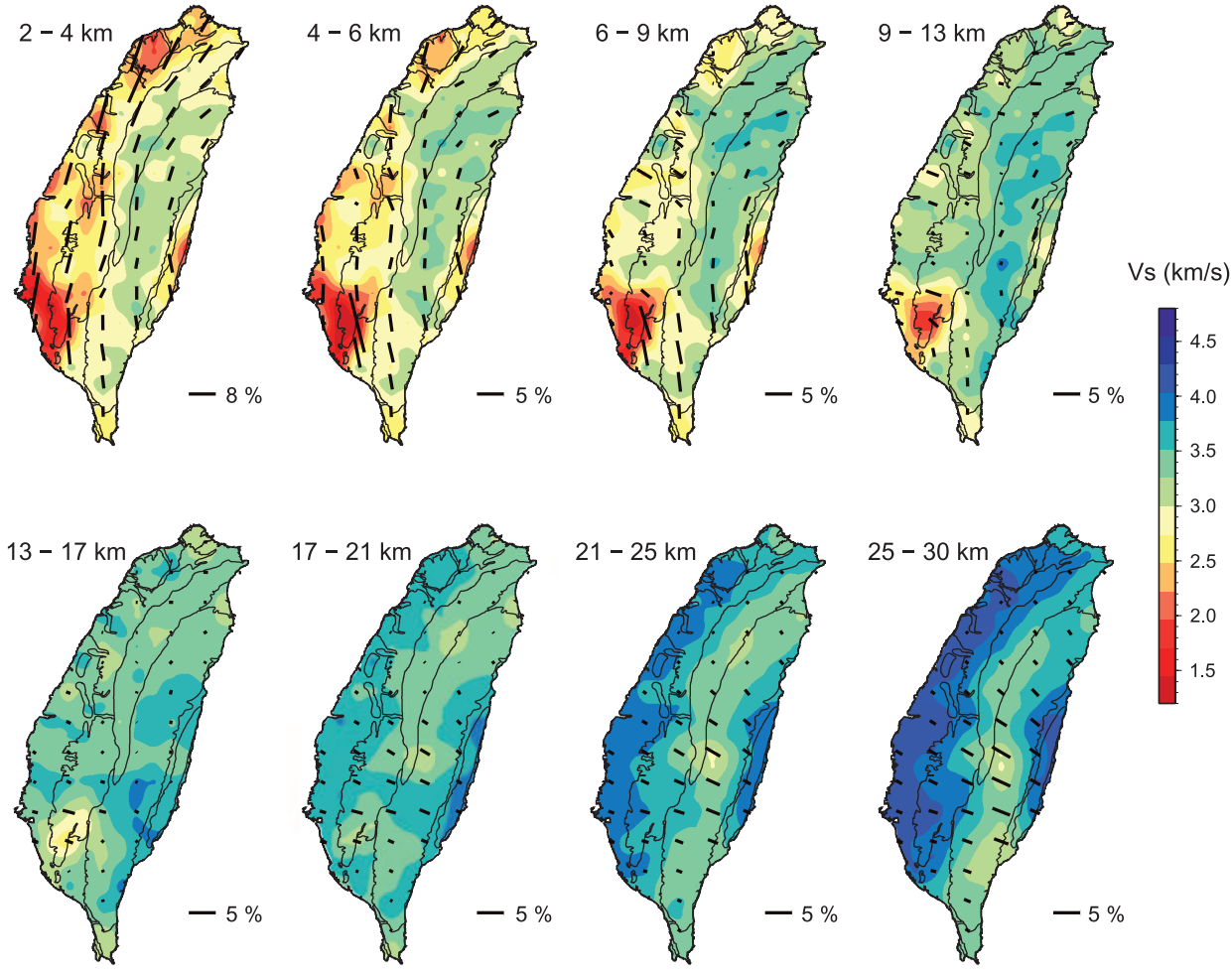


Fig. 2. Maps of *Vs* and azimuthal anisotropy at selected depths. The *Vs* in the high mountains is relatively high in the upper crust but relatively low in the lower crust. The fast direction of azimuthal anisotropy is represented by the orientation of the bar with its length proportional to the strength of anisotropy indicated in the lower right of each panel. The fast directions rotate from nearly orogen-parallel to nearly convergence-parallel across 10- to 15-km depths.

of the CPA requires shearing to occur mostly within the subducted Eurasian crust, where relative motion can be imposed between the surface of the subducted plate, where collision exerts resistance, and the continuously sinking mantle lithosphere (Fig. 4). The CPA diminishes rapidly toward northern Taiwan (Fig. 2), where subduction of the EP is absent, in agreement with the proposed subduction-induced mechanism.

The proposed shear deformation beneath the HR and CR may occur in the regime of ductile flow. In this region, with estimated temperatures (~750°C at a 25-km depth) (*20*) comparable to those suggested for the ductile lower crust of Tibet (*3*), the seismicity is considerably reduced relative to the surroundings. Below the CR, the relatively low *Vs* in the CPA zone is compatible with elevated temperatures. The magnitude of the CPA increases with depth to 30 km, implying an enhanced shear flow with increasing temperature. The viscosity below the CR at depths corresponding to the CPA zone is lower than 10^{19} from postseismic deformation (*21*), which is lower than required to maintain ductile flow in Tibet (*3*). Toward the west, the change in mineralogy at shallower depth and the elevation of the CPA zone to the brittle deformation regime may reduce the strength of the anisotropy.

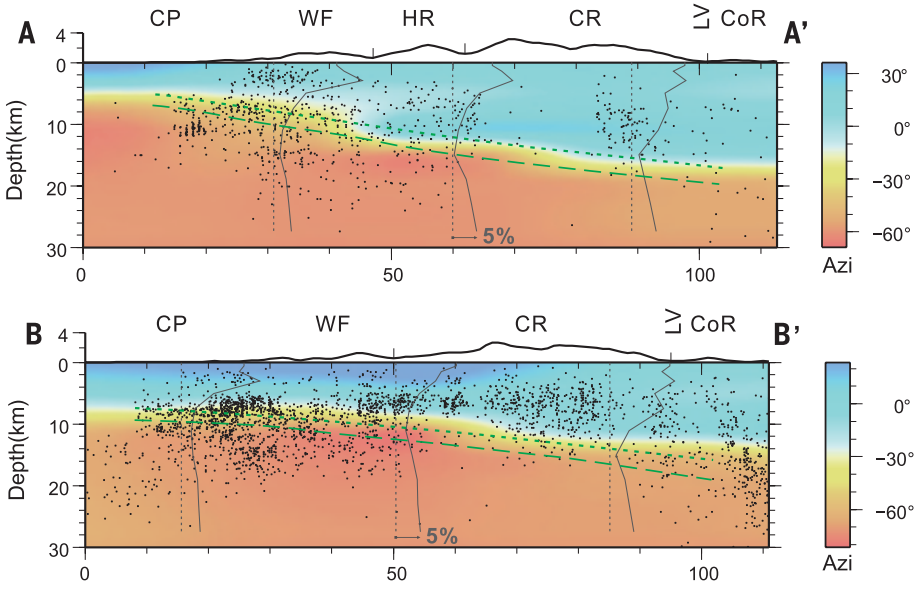


Fig. 3. Cross sections showing rotation of anisotropy fabrics with depth. The positions of AA′ and BB′ are shown in Fig. 1. The resolved fast directions of azimuthal anisotropy are color coded to highlight the rotation from 30° to 0°, to −60° to −90° over a narrow depth range. Green dotted lines represent the parameterized ATB, which in general agrees with the color description of the rotation. Green long dashed lines mark the top of the CPA zone. Both sections are decorated by seismicity during 2003 to 2007. In each section, three representative profiles of the strength of anisotropy are shown in gray lines, and the reference for zero amplitude is shown by the vertical dotted lines.

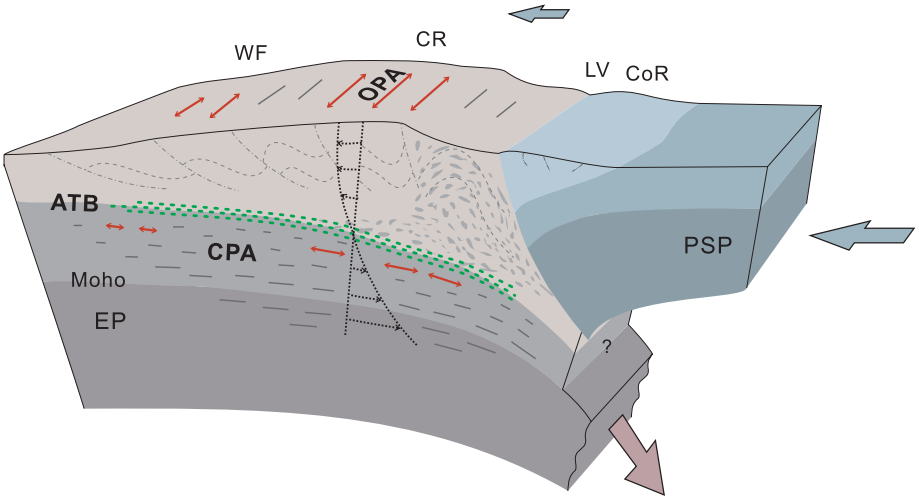


Fig. 4. Cartoon illustration of the layered deformation zones in the Taiwan orogen. The compressional tectonics and the subduction-dominated deformation are characterized by the OPA and the CPA, respectively. Red double arrows highlight the fast directions of these two sets of anisotropy. The ATB (green dots) separates the two deformation regimes and couples the upper crust of the orogen with the subduction. Light blue arrows indicate the motion of the PSP and the orogen. The brown arrow indicates the continuous sinking of the mantle lithosphere of the EP while the upper part of the plate is coupled with the orogen. The schematic velocity profile illustrates the relative motion of the collision-subduction system. The exhumation of the continental basement material beneath the CR is illustrated with irregular spindles.

The ATB averaged in central Taiwan dips more steeply (8.3°) than that in southern Taiwan (4.8°) in the range between the CP and the CR. Below the CR, the ATB is on average 5 km deeper in central than in southern Taiwan. This is compatible with the geochemical evidence that the pre-Tertiary metamorphic rocks exposed in the eastern CR (the Tananao Complex) were exhumed from the upper crust of the underthrusted continental margin (*22*). The ATB zone may initiate in the vicinity of the surface of the EP but carve into the underthrusted crust as the latter is peeled off during collision. The maximum depth of exhumation beneath the eastern CR may be constrained at the top of the CPA, at roughly a 20-km depth, where shearing fully develops. This adds a constraint on the models that incorporate deep exhumation (*23*, *24*). In southern Taiwan, where orogeny might have started but not culminated yet (*25*) and no Eurasian basement outcrops have been found (*26*), the ATB retains a smaller slope and is shallower beneath the CR.

A previous hypothesis suggested a décollement separating the upper crust from the underthrusted plate (*5*). The ATB interpreted here is related to a no-slip boundary condition between the upper and lower crustal levels. Because of this coupling, the shear zone extends into the crust and probably the mantle of the down-going plate, in contrast to the focused shear and slip along a single detachment. Some portion of the convergence may be taken up by the ductile flow across this broad shear zone. However, the layered deformation implies that the compressional tectonics is confined above the ATB, which echoes the central principle of the thin-skinned tectonics.

In the sense of the distribution of deformation, our observation is consistent with the thick-skinned tectonics (*7*–*19*). However, previous studies supportive of thick-skinned model emphasized (i) a continuous deformation to counter the concept of the décollement and (ii) a large-scale compressional deformation dominated by collision. We argue that the collision produces compressional tectonics in the upper crust of the orogen but, together with the sinking slab, induces shear within the lower crust or the crust of the subducted EP, rather than the entire crust being shortened and thickened. We speculate that the shear is an ambient state of stress of the subducted plate, which may not be easily overprinted

by the deep-penetrating faults (*8*). Our interpretation suggests that the ongoing formation of the mountain of Taiwan is coupled with the subduction through shearing. The coupling might be spatially heterogeneous and evolve in time as subduction progresses to collision.

REFERENCES AND NOTES

1. V. Schulte-Pelkum *et al*., *Nature* **435**, 1222–1225 (2005).
2. J. Nábelek *et al*., *Science* **325**, 1371–1374 (2009).
3. C. Beaumont, R. A. Jamieson, M. H. Nguyen, S. Medvedev, *J. Geophys. Res.* **109**, B06406 (2004).
4. T. Byrne *et al*., The arc-continent collision in Taiwan, in *Arc-Continent Collision, Frontiers in Earth Sciences*, D. Brown, P. Ryan, Eds. (Springer-Verlag, Berlin, 2011), pp. 213–245.
5. D. Davis, J. Suppe, F. A. Dahlen, *J. Geophys. Res.* **88**, 1153–1172 (1983).
6. S. Carena, J. Suppe, H. Kao, *Geology* **30**, 935 (2002).
7. F. Wu, R.-J. Rau, D. Salzberg, *Tectonophysics* **274**, 191–220 (1997).
8. E. A. Bertrand *et al*., *J. Geophys. Res.* **117**, B01402 (2012).
9. F. T. Wu, H. Kuo-Chen, K. D. McIntosh, *J. Asian Earth Sci.* **90**, 173–208 (2014).
10. P. G. Silver, *Annu. Rev. Earth Planet. Sci.* **24**, 385–432 (1996).
11. H. Kuo-Chen, F. T. Wu, D. Okaya, B.-S. Huang, W.-T. Liang, *Geophys. Res. Lett.* **36**, L12303 (2009).
12. W. Huang *et al*., *J. Geophys. Res.* **105** (B12), 27979–27989 (2000).
13. T.-Y. Huang, Y. Gung, W.-T. Liang, L.-Y. Chiao, L. S. Teng, *Geophys. Res. Lett.* **39**, L05305 (2012).
14. L.-Y. Chiao, B.-Y. Kuo, *Geophys. J. Int.* **145**, 517–527 (2001).
15. Materials and methods are available as supplementary materials on *Science* Online
16. C.-H. Chen, Y.-H. Chen, H.-Y. Yen, G.-K. Yu, *Earth Planets Space* **55**, 223–230 (2003).
17. R. L. Rudnick, S. Gao, Composition of the continental crust, in *Treatise of Geochemistry 2nd ed.*, H. Holland, K. Turekian, Eds. (Elsevier, Amsterdam, 2014).
18. D. J. Tatham, G. E. Lloyd, R. W. H. Butler, M. Casey, *Earth Planet. Sci. Lett.* **267**, 118–128 (2008).
19. D. Mainprice, A. Nicolas, *J. Struct. Geol.* **11**, 175–189 (1989).
20. H. Kuo-Chen *et al*., *Geophys. Res. Lett.* **39**, L22302 (2012).
21. Y.-J. Hsu, S.-B. Yu, M. Simons, L.-C. Kuo, H.-Y. Chen, *Tectonophysics* **479**, 4–18 (2009).
22. W. G. Ernst, *J. Metamorph. Geol.* **1**, 305–329 (1983).
23. M. Simoes *et al*., *J. Geophys. Res.* **112**, B11405 (2007).
24. P. Yamato, F. Mouthereau, E. Burov, *Geophys. J. Int.* **176**, 307–326 (2009).
25. Y.-H. Lee *et al*., *Geology* **43**, 451–454 (2015).
26. C.-S. Ho, *An Introduction to the Geology of Taiwan* (Central Geological Survey, Taiwan, 1999).
27. B. Rousset, S. Barbot, J.-P. Avouac, Y.-J. Hsu, *J. Geophys. Res.* **117**, B12405 (2012).

ACKNOWLEDGMENTS

We thank W.-T. Liang, F. T. Wu, K.-L. Wang, T.-F. Yui, W.-S. Chen, and J. Suppe for helpful discussions and comments. T. Byrne helped clarify the geological implications in an early draft and provided inspiring discussion. We thank three reviewers for their critical comments. We acknowledge the Taiwan Integrated Geodynamics Research (TAIGER) project, The Taiwan Earthquake Research Data Center (TECDC), Broad-Band Array in Taiwan for Seismology (BATS), and Central Weather Bureau Broad-Band array (CWBBB) for providing continuous broadband seismic waveforms. This work was supported by the Ministry of Science and Technology of Taiwan (MOST 103-2116-M-002 -026) and Academia Sinica, Taiwan (AS-104-TP-A04). All the data used in this study are available in the following websites: CWBBB: http://gdms.cwb.gov.tw/index.php; BATS: http://bats.earth.sinica.edu.tw; TAIGER: http://ds.iris.edu/mda/YM?timewindow=2006-2009.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/720/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 and S2
References (*28*–*35*)

24 March 2015; accepted 30 June 2015
10.1126/science.aab1879

2D MATERIALS

Observation of tunable band gap and anisotropic Dirac semimetal state in black phosphorus

Jimin Kim,[1] Seung Su Baik,[2,3] Sae Hee Ryu,[1,4] Yeongsup Sohn,[1,4] Soohyung Park,[2] Byeong-Gyu Park,[5] Jonathan Denlinger,[6] Yeonjin Yi,[2] Hyoung Joon Choi,[2,3] Keun Su Kim[1,4]*

Black phosphorus consists of stacked layers of phosphorene, a two-dimensional semiconductor with promising device characteristics. We report the realization of a widely tunable band gap in few-layer black phosphorus doped with potassium using an in situ surface doping technique. Through band structure measurements and calculations, we demonstrate that a vertical electric field from dopants modulates the band gap, owing to the giant Stark effect, and tunes the material from a moderate-gap semiconductor to a band-inverted semimetal. At the critical field of this band inversion, the material becomes a Dirac semimetal with anisotropic dispersion, linear in armchair and quadratic in zigzag directions. The tunable band structure of black phosphorus may allow great flexibility in design and optimization of electronic and optoelectronic devices.

Two-dimensional (2D) atomic crystals have continued to show great potential for application in nanoscale devices (*1*). A key issue is controlling their electronic states to overcome the limit of natural properties. Graphene's effectively massless state of charge carriers is a result of the conical band structure (*2*). However, the lack of a band gap (E_g) limits the on-off current ratio in planar field-effect transistors (*3*, *4*), and it has been difficult to reliably achieve a sizable E_g without degrading its electronic quality (*5*–*7*). It would thus be desirable to realize a 2D system with a widely tunable E_g.

A potential candidate is few-layer black phosphorus (BP), a layered material of elemental phosphorus (*5*–*22*). The single-layer BP (or phosphorene) has a honeycomb network similar to graphene but is strongly puckered (armchair-shaped along x and zigzag-shaped along y in Fig. 1A), rendering its electronic state highly susceptible to external perturbations (*11*–*22*). The low-energy band structure of phosphorene can be approximated by a bonding and antibonding pair of mainly $3p_z$ orbitals (*11*, *12*). The corresponding valence band (VB) and conduction band (CB) are located at the zone center (Γ_2^+ and Γ_4^- states in Fig. 1B) with the predicted E_g of 0.7 to 1.6 eV (*13*–*17*). For multilayers, the introduction of interlayer coupling reduces E_g with increasing film thickness and reaches ~0.33 eV in bulk BP (*11*–*14*). The E_g of BP films and nanoribbons has been widely predicted to be tunable by strain (*15*–*17*) and electric field (*17*–*21*), the latter of which is more viable in gated devices. The electric field affects the real-space distribution of VB and CB states to be shifted in opposite directions to each other. Their potential difference and band mixing lead to a reduction in E_g, which is often termed the giant Stark effect (*23*–*26*). However, despite its potential importance for device applications, little is known experimentally about this effect on the surface of 2D semiconductors under a vertical electric field.

Here we report the realization of a widely tunable E_g in BP by means of the in situ deposition of potassium (K) atoms, the well-known technique to induce doping and electric field in 2D van der Waals systems (*27*). The K atoms on BP donate charges to the few top phosphorene layers, which are confined to form 2D electron gas near the surface (Fig. 1A, bottom). The band structure measured by angle-resolved photoemission spectroscopy (ARPES) at low K density is slightly n-doped with E_g greater than 0.6 eV (Fig. 1C). With increasing dopant density, the electric field from the ionized K donors gradually reduces E_g owing to the giant Stark effect, as supported by our density functional theory (DFT) calculations. Consequently, the electronic state of BP is widely and continuously tuned from a moderate-gap semiconductor to a band-inverted semimetal. At the critical dopant density of this band-inverted transition (*21*, *22*), the electronic state becomes an anisotropic Dirac semimetal (Fig. 1D). This control mechanism of E_g should work in dual-gate BP devices for precisely balancing between high mobility and moderate E_g.

Figure 1, E to H, shows a series of ARPES spectra taken along the armchair direction k_x with different dopant density near direct E_g. As expected for pristine BP (*28*, *29*), in Fig. 1E there

[1]Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Korea. [2]Department of Physics, Yonsei University, Seoul 120-749, Korea. [3]Center for Computational Studies of Advanced Electronic Material Properties, Yonsei University, Seoul 120-749, Korea. [4]Center for Artificial Low Dimensional Electronic Systems, Institute for Basic Science, Pohang 790-784, Korea. [5]Pohang Accelerator Laboratory, Pohang University of Science and Technology, Pohang 790-784, Korea. [6]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
***Corresponding author. E-mail: keunsukim@postech.edu**

is a well-defined VB with a nearly parabolic dispersion and with a vertex at 0.15 eV below the Fermi energy E_F (*30*). Assuming the reported E_g of 0.33 eV (*12*), our pristine sample is slightly hole-doped, which explains p-type conduction in transport (*12*). We use a k·p perturbation formula (*31*), widely accepted to fit the band dispersion of narrow-gap semiconductors, to quantify the hole effective mass m_x^*. The best fit (white line overlaid) yields $m_x^* = 0.08 \pm 0.03 m_e$, where m_e is the electron rest mass, in good agreement with theoretical calculations (*12*, *14*).

Upon electron doping by the K deposition on the surface, the overall band structure rigidly shifts down toward high energies (fig. S2), so that the magnitude of E_g can be directly measured by ARPES. The energy shift of surface bands is accompanied by steep band bending toward the bulk (along z in Fig. 1A) to form 2D electronic states in a few phosphorene layers, as confirmed by little k_z dependence (*31*). At low dopant density just before the CB minimum drops below E_F (Fig. 1F), the VB rigidly shifts down, and E_g can be estimated at ~0.6 eV or slightly greater. This E_g of 2D electronic states is smaller than that predicted for monolayer phosphorene (*13*–*17*) and is comparable to those predicted for few-layer phosphorene (0.3 to 1.0 eV) (*13*, *14*). With further increasing dopant density, this E_g gradually reduces (Fig. 1G) and becomes zero (Fig. 1H), which indicates a semiconductor-semimetal transition. At the critical dopant density for this transition, where the VB and CB touch each other, the band dispersion in the armchair direction k_x becomes linear (Fig. 1H), whereas that in the zigzag direction k_y remains nearly parabolic (fig. S3).

The linear dispersion in k_x can be more clearly identified in high-resolution ARPES data (Fig. 2A). The peak positions follow X-shaped linear bands of the VB (red) and CB (blue) with a crossing point (Fig. 2C) that is also revealed in the intensity map at the crossing energy (Fig. 2E). The spectral simulation with linear bands and finite broadening (*31*) (Fig. 2B) successfully reproduces the experimental data in Fig. 2A. This suggests that BP at the critical density has a spin-degenerate Dirac state as in graphene (*3*) and Dirac semimetals (*21*, *32*–*34*), which can be stabilized by crystalline symmetry, as explained below. A linear fit yields the velocity of charge carriers $v_x = 5.1 \pm 0.9 \times 10^5$ m/s, which is about half of that in graphene (*2*). Fig. 2, D to G, shows a series of constant-energy intensity maps. The ARPES intensity of band contours is modulated by the matrix-element effect that reflects the symmetry of constituent wave functions in real space (*5*). The maps above and below the energy of the Dirac point E_D (Fig. 2, D and F) show strong suppression along k_y with respect to k_x. The resultant twofold symmetry of intensity patterns confirms that they originate from BP rather than from K. Taking this matrix-element effect into account, the band contours are oval-shaped with the aspect ratio $r \sim 1.9$, which is related to armchair-zigzag anisotropy in Fig. 1A. Around the contour, the band dispersion gradually changes from linear along k_x to quadratic along k_y (fig. S4). These results support the picture of the anisotropic band crossing (Fig. 1D),

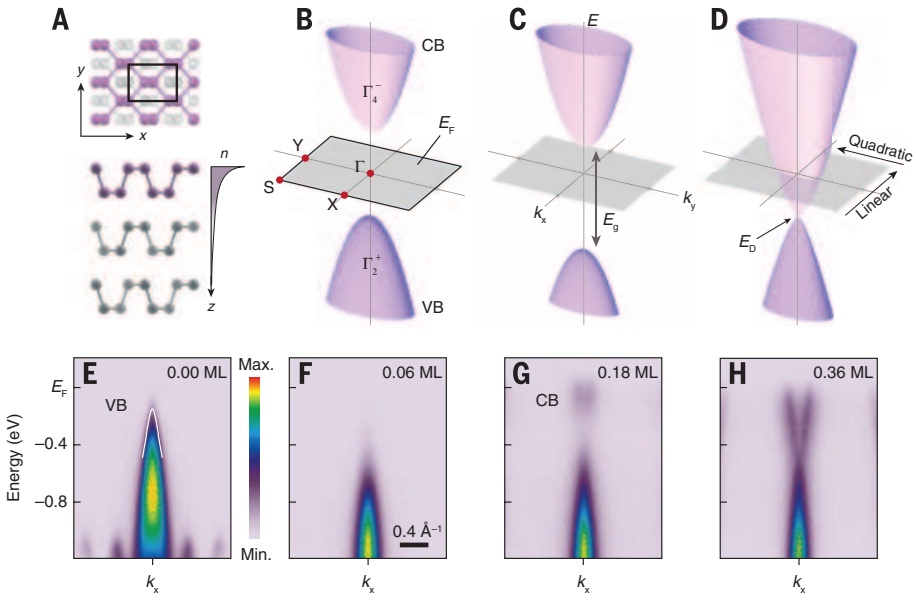


Fig. 1. Band structure and tunable E_g of few-layer BP. (**A**) (Top) Atomic structure of BP. The solid square is the surface unit cell, and the interlayer distance is ~5 Å. (Bottom) Density profile of a 2D electron gas decaying along z toward the bulk. (**B** to **D**) Band structure of (B) pristine BP, (C) BP in the initial stage of surface doping, and (D) BP at the transition to a zero-gap semimetal. The solid square in (B) indicates the surface Brillouin zone with high symmetry points marked by red circles. (**E** to **H**) Experimental band structure of BP taken at 15 K near E_F along k_x with dopant density marked at the upper right of each panel. The photon energy is 104 eV for k_z at the Z point of the bulk Brillouin zone (*29*). The dopant density is estimated in units of monolayers from simultaneously taken K 3p core-level spectra (fig. S7). The white line overlaid in (E) is a fit to the VB with the k·p perturbation formula (*31*).

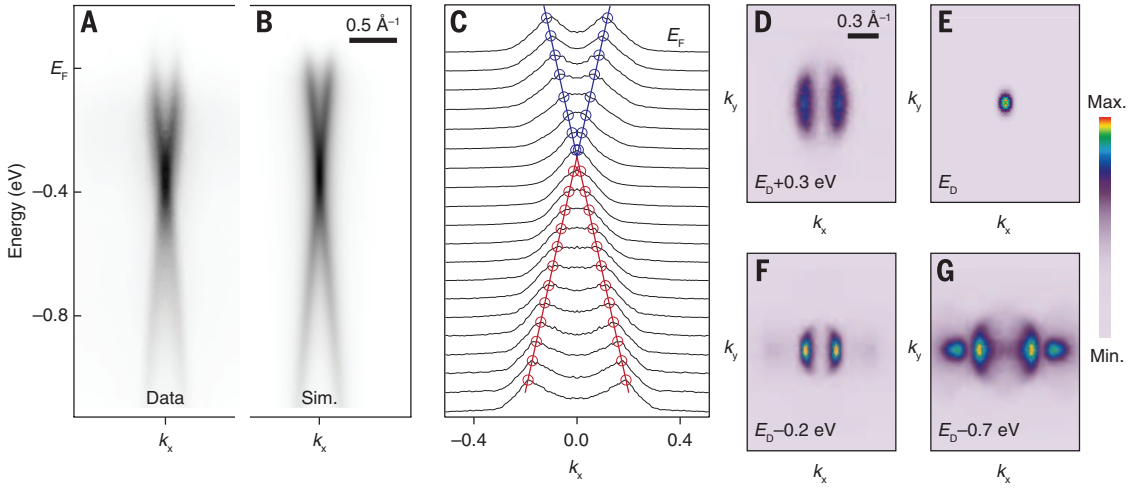


Fig. 2. Anisotropic Dirac semimetal state at the critical density. (**A**) High-resolution ARPES data taken at 15 K along the armchair direction k_x. (**B**) Corresponding spectral simulation with two linearly crossing bands and finite broadening (*31*). (**C**) Normalized momentum-distribution curves (0.05-eV steps from E_F) with their peak positions marked by open circles. Red and blue lines are linear fits to the VB and CB, respectively. (**D** to **G**) Series of ARPES intensity maps at constant energies (marked at the bottom of each panel), shown over a whole surface Brillouin zone (Fig. 1B).

which is reproduced by DFT calculations, as shown below.

To systematically follow the evolution of the band structure, we display a 3D representation of ARPES spectra as a function of dopant density N in Fig. 3A. The constant-energy cut at E_F shows that the Fermi momentum k_F of the CB steadily increases. We estimate the electron concentration n based on Luttinger's theorem as $\pi k_F^2 r$ that corresponds to the area enclosed by oval-shaped contours at E_F. We found that n is linearly proportional to N (fig. S6), which indicates monotonic charge transfer from dopants to phosphorene layers. In the constant-momentum cut at $k_x = 0$, diamonds and circles denote the positions of the VB maximum (E_v) and CB minimum (E_c), respectively. At the initial stage of doping, the Γ_2^+ state at E_v has rapidly shifted down until the Γ_4^- state at E_c drops below E_F. As N is increased further, the center energy $E_0 = (E_v + E_c)/2$ slowly shifts down, whereas the Γ_2^+ and Γ_4^- states get progressively closer to each other and eventually cross at the critical density $N_c = 0.36$ monolayers (ML) ($n = 8.3 \times 10^{13}$ cm^{-2}), where band inversion occurs. The magnitude of E_g is calculated as $E_c - E_v$ and is plotted as a function of N in Fig. 3B. As can be seen from the figure, E_g is widely and continuously tunable in the range of +0.6 to −0.2 eV.

We performed DFT calculations based on four-layer BP with a single K atom on each 2-by-2 surface unit cell (corresponding to $N \sim 0.4$ ML in experiments). To effectively describe a lower K density, we increase the vertical distance d between K and BP (2.76 Å at equilibrium) such that the effect of K donors is gradually reduced without change in the supercell size (*31*). A series of band calculations as a function of d reproduces key aspects of our experimental observations, the variation of E_g and resultant semimetal-semiconductor transition (red circles in Fig. 3B). Furthermore, at $E_g \sim 0$ ($d = 3.67$ Å), the bands along k_x are linearly dispersing near E_D, whereas those along k_y remain parabolic (Fig. 3C), as observed experimentally. N_c in Fig. 3B corresponds to the critical point of the band-inverted transition, where the topological invariant quantity Z_2 switches between 0 and 1 (*21*), induced by electric field rather than spin-orbit interaction. At this quantum critical point, the system is predicted to be in an unusual Dirac semimetal state (*32*) in which the band crossing along the rotational symmetry axis (the y axis, zigzag) is quadratic, whereas that along the other axis (the x axis, armchair) is linear. Our results thus collectively identify the formation of the anisotropic Dirac semimetal state at N_c, resulting from the characteristic puckered structure of phosphorene.

We now discuss the control mechanism of E_g. Figure 3D shows partial charge densities of E_v and E_c points in Fig. 3C, separated by a tiny gap (less than 10 meV) to avoid their degeneracy. The spatial distributions of E_v and E_c states, which are uniform in pristine BP (fig. S9), become strongly separated in opposite directions, indicating an electric field generated by the ionized K donors. The E_c states, which have a positive effective mass along z, freely move toward the positive electrical potential, whereas the E_v states, which have a negative effective mass along z, are pushed within the body of the BP layers (*12*). As stated above, this marked spatial separation of E_v and E_c states explains the variation of E_g by the giant Stark effect (*17*–*21*). Band calculations for four-layer BP under external electric field confirm the similar reduction in E_g (with no change in E_F) and the band crossover at the critical field of ~0.19 V/Å. From this value, we quantify the Stark coefficient S_L for four-layer BP as ~3 Å (*24*), which is comparable to those predicted for few-layer BP (*19*, *21*) and transition metal dichalcogenides (*25*, *26*). Because S_L is known to increase with film thickness (*19*, *21*), the critical field for bulk BP (or thicker BP films) would be smaller than the practical dielectric strengths of SiO_2 and h-BN. Our work thus demonstrates the giant Stark effect in BP as an efficient control mechanism of E_g, which is generally attainable in 2D semiconductors and devices composed of these materials.

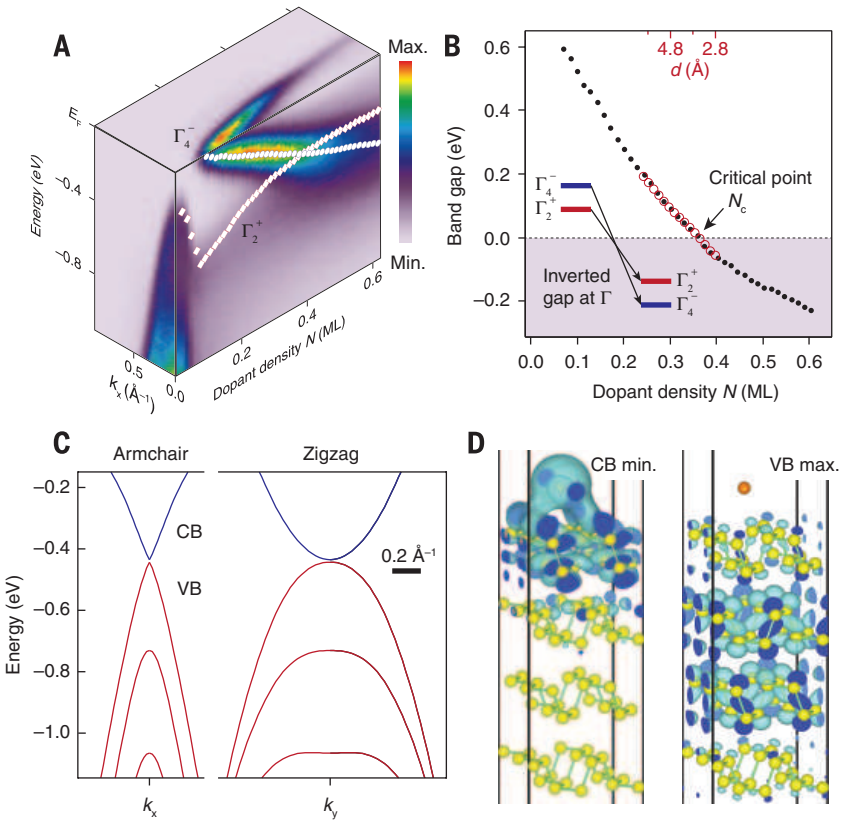


Fig. 3. Quantitative analysis for the control mechanism of E_g. (**A**) A 3D representation of the band evolution as a function of N, taken at 15 K. Diamonds and circles denote the positions of E_v and E_c, respectively. (**B**) E_g as a function of N. The negative gap in the shaded region represents the inverted gap at the Γ point. Overlaid red circles denote E_g from DFT band calculations where varying dopant density is simulated by changing the vertical distance d between K and BP (*31*). (**C**) Theoretical band dispersions for K-doped BP films at $d = 3.67$ Å. (**D**) Partial charge densities of E_v and E_c points in (C), separated by a tiny gap of less than 10 meV. Light blue areas represent the isosurface, set to 1.76 × 10^{-3}/Å^3, and its cross sections are shown in dark blue. Yellow and orange balls represent P and K atoms, respectively.

REFERENCES AND NOTES

1. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
2. K. S. Novoselov *et al.*, *Nature* **438**, 197–200 (2005).
3. A. K. Geim, *Science* **324**, 1530–1534 (2009).
4. L. Britnell *et al.*, *Science* **335**, 947–950 (2012).
5. K. S. Kim *et al.*, *Nat. Mater.* **12**, 887–892 (2013).
6. E. S. Reich, *Nature* **506**, 19 (2014).
7. H. O. H. Churchill, P. Jarillo-Herrero, *Nat. Nanotechnol.* **9**, 330–331 (2014).
8. L. Li *et al.*, *Nat. Nanotechnol.* **9**, 372–377 (2014).
9. H. Liu *et al.*, *ACS Nano* **8**, 4033–4041 (2014).
10. S. P. Koenig, R. A. Doganov, H. Schmidt, A. H. Castro Neto, B. Özyilmaz, *Appl. Phys. Lett.* **104**, 103106 (2014).
11. Y. Takao, H. Asahina, A. Morita, *J. Phys. Soc. Jpn.* **50**, 3362–3369 (1981).
12. A. Morita, *Appl. Phys.* **39**, 227–242 (1986).
13. A. N. Rudenko, M. I. Katsnelson, *Phys. Rev. B* **89**, 201408 (2014).
14. J. Qiao, X. Kong, Z.-X. Hu, F. Yang, W. Ji, *Nat. Commun.* **5**, 4475 (2014).
15. A. S. Rodin, A. Carvalho, A. H. Castro Neto, *Phys. Rev. Lett.* **112**, 176801 (2014).
16. X. Peng, Q. Wei, A. Copple, *Phys. Rev. B* **90**, 085402 (2014).
17. Y. Li, S. Yang, J. Li, *J. Phys. Chem. C* **118**, 23970–23976 (2014).
18. J. Dai, X. C. Zeng, *J. Phys. Chem. Lett.* **5**, 1289–1293 (2014).
19. H. Guo, N. Lu, J. Dai, X. Wu, X. C. Zeng, *J. Phys. Chem. C* **118**, 14051–14059 (2014).
20. Q. Wu, L. Shen, M. Yang, Z. Huang, Y. P. Feng, http://arxiv.org/abs/1405.3077.
21. Q. Liu, X. Zhang, L. B. Abdalla, A. Fazzio, A. Zunger, *Nano Lett.* **15**, 1222–1228 (2015).
22. R. Fei, V. Tran, L. Yang, *Phys. Rev. B* **91**, 195319 (2015).
23. K. H. Khoo, M. S. C. Mazzoni, S. G. Louie, *Phys. Rev. B* **69**, 201401 (2004).
24. F. Zheng, Z. Liu, J. Wu, W. Duan, B.-L. Gu, *Phys. Rev. B* **78**, 085423 (2008).

25. A. Ramasubramaniam, D. Naveh, E. Towe, *Phys. Rev. B* **84**, 205325 (2011).
26. Q. Yue *et al.*, *J. Phys. Condens. Matter* **24**, 335501 (2012).
27. T. Ohta, A. Bostwick, T. Seyller, K. Horn, E. Rotenberg, *Science* **313**, 951–954 (2006).
28. T. Takahashi, H. Tokailin, S. Suzuki, T. Sagawa, I. Shirotani, *Phys. Rev. B* **29**, 1105–1108 (1984).
29. C. Q. Han *et al.*, *Phys. Rev. B* **90**, 085101 (2014).
30. Note that continuum-like features in VBs vary with samples (see fig. S3), probably due to natural stacking disorder (*5*). Accordingly, the relative intensity of VBs and CBs is different for each cleaving of samples.
31. Materials and methods are available as supplementary materials on *Science* Online.
32. B.-J. Yang, N. Nagaosa, *Nat. Commun.* **5**, 4898 (2014).
33. Z. K. Liu *et al.*, *Science* **343**, 864–867 (2014).
34. X. Qian, J. Liu, L. Fu, J. Li, *Science* **346**, 1344–1347 (2014).

ACKNOWLEDGMENTS

This work was supported by IBS-R014-D1. S.S.B. and H.J.C. acknowledge support from the National Research Foundation of Korea (grant 2011-0018306). Computational resources have been provided by Korea Institute of Science and Technology Information Supercomputing Center (project KSC-2013-C3-062). S.P. and Y.Y. acknowledge support from the National Research Foundation of Korea (grant 2013R1A1A1004778) and Yonsei University Future-Leading Research Initiative of 2014 (2014-22-0123). PLS-II was supported in part by the Ministry of Science, ICT and Future Planning and Pohang University of Science and Technology. The Advanced Light Source was supported by the U.S. Department of Energy, Office of Sciences, under contract DE-AC02-05CH11231. We thank S. W. Jung, W. J. Shin, Y. K. Kim, and B. Kim for help with ARPES experiments.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/723/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*35*–*41*)

8 January 2015; accepted 14 July 2015
10.1126/science.aaa6486

APPLIED OPTICS

Voltage-tunable circular photogalvanic effect in silicon nanowires

Sajal Dhara,[1] Eugene J. Mele,[2] Ritesh Agarwal[1]*

Electronic bands in crystals can support nontrivial topological textures arising from spin-orbit interactions, but purely orbital mechanisms can realize closely related dynamics without breaking spin degeneracies, opening up applications in materials containing only light elements. One such application is the circular photogalvanic effect (CPGE), which is the generation of photocurrents whose magnitude and polarity depend on the chirality of optical excitation. We show that the CPGE can arise from interband transitions at the metal contacts to silicon nanowires, where inversion symmetry is locally broken by an electric field. Bias voltage that modulates this field further controls the sign and magnitude of the CPGE. The generation of chirality-dependent photocurrents in silicon with a purely orbital-based mechanism will enable new functionalities in silicon that can be integrated with conventional electronics.

In the circular photogalvanic effect (CPGE), the polarity and magnitude of photocurrents can be controlled by the chirality of elliptically polarized optical excitation (P_{circ}) in a certain class of materials known as gyrotropic media. The effect originates from the unequal population of excited charged carriers in a preferential momentum direction when excited by light with left ($\sigma = -1$) or right ($\sigma = +1$) circular polarization. Semiconductors that support the CPGE are traditionally gyrotropic optical media with a strong spin-orbit coupling, so that the effect is ordinarily controlled by angular momentum selection rules for excitation with circularly polarized light. The effect has been observed in different quantum-well (QW) structures (*1*–*5*), for which it is attributed to the *k*-linear spin splitting of energy bands due to the spin-orbit interaction.

Most materials that have found widespread applications in conventional electronics (e.g., centrosymmetric crystals such as Si and Ge) are not gyrotropic and do not exhibit a bulk CPGE unless quantum-confined to below 10-nm length scales. In Si/Ge QWs, a CPGE due to orbital (*6*, *7*) or valley-orbital interactions (*8*), appears in the long wavelength range (~100 μm). The underlying mechanism is attributed to intrasubband free carrier absorption pathways interfering with intersubband excitation (*6*) via polarizability effects. Photogalvanic effects have also been theoretically predicted in carbon nanotubes, without involving the electron spin degree of freedom (*9*). Because there are similarities between chiral nanotubes and Si nanowires (NWs) (*10*–*12*), it is desirable to determine whether centrosymmetric crystals of technologically important materials such as Si can also exhibit a CPGE when their shapes are engineered, thus adding novel functionalities.

We demonstrated a mechanism for the CPGE involving only the orbital degrees of freedom that are observed at the surface of Si NWs at the metal-NW junction. The atomic structure of the NW, along with the macroscopic field present at the contact, breaks the bulk symmetries allowing the CPGE. The $[1\bar{1}0]$surface of Si is of particular interest because of its high hole mobility (*13*–*15*) associated with a zigzag chain of atoms running along the ⟨110⟩ direction. A Schottky electric field along the NW ⟨111⟩ growth direction breaks the relevant mirror symmetries and produces a chiral structure, producing a CPGE that is tunable with applied bias.

Unintentionally doped Si NWs (diameter, 100 to 150 nm; grown mostly along the ⟨111⟩ direction and some along the ⟨112⟩ direction) were used to make two-terminal devices by electron-beam lithography [section 1 in (*16*)]. Figure 1A shows the schematic of the device and the measurement setup, where the laser [transverse electromagnetic (TEM_{00}) mode, wavelength 680 nm, spot size ~2 μm] is incident at the metal-NW junction [see (*16*) for details in section 1 and figs. S1 to S6]. The excitation energy (680 nm) of the laser is above the indirect bandgap of Si, and therefore contributions to the photocurrent from interband excitation (corresponding to the bulk of the Si NW) and the surface states on the $[1\bar{1}0]$ plane can be simultaneously present, albeit with different selection rules. Because of the geometrical anisotropy of NWs, the bulk contribution to photocurrent has two maxima with respect to linear polarization, owing to the preferential absorption of light polarized along the long axis (TM polarization) (*17*, *18*) and in the perpendicular direction (TE polarization) to the NW at the metal-NW junction due to metal antenna effects (*19*, *20*). In all of our experiments, we observed the usual linear polarization–dependent photocurrent variation as a function of the quarter-wave plate (QWP) [half-wave plate angle–dependent photocurrent variation is discussed in (*16*), sections 2 and 3 and fig. S8]. However, at the metal-NW junction, the most important contribution to the photocurrent came from the P_{circ} dependence, which we observed by noticing that the maximum change in the photocurrent appeared at the left-handed (45°) and right-handed (135°) circular polarization, which has a variation of $\sin(2\alpha)$ irrespective of the physical orientation of the NW device [(*16*), section 4 and fig. S9]. As a result, the observed photocurrent at room temperature of a representative device at an applied bias of 1 V at two different spatial positions, in the metal-NW contact region (Fig. 1B), and on the bare NW (far away from the contacts, Fig. 1C) displayed different polarization dependence. At the metal-semiconductor contact, the photocurrent pattern repeated twice [$\propto \sin(2\alpha)$] as the QWP angle was changed from 0° to 360°, whereas on the NW, the pattern repeated four times [$\sim\cos(4\alpha + \phi)$, with a phase term (ϕ) depending on the physical orientation of the NW and laser polarization], suggesting a strong P_{circ}–dependent photocurrent confined to the contact region. The photocurrents can be fitted (Fig. 1, B and C, solid line) with the expression $I(\alpha) = I_c \sin 2\alpha + I_l \cos(4\alpha + \varphi) + I_d$, where I_c, I_l, and I_d are

[1]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA.
***Corresponding author. E-mail: riteshag@seas.upenn.edu**

the coefficients for circular polarization (CP), linear polarization (LP), and polarization-independent components of the photocurrent, respectively. Figure 1B shows that the CP-dependent photocurrent (normalized with laser intensity) was observed with a value of I_c = 1.0 nA kW^{-1} cm^2 at the metal-NW junction, and it reversed its sign with respect to a background current, I_d = 5.2 nA kW^{-1} cm^2, at α = 45° (left circular polarization) and 135° (right circular polarization). The LP-dependent contribution, I_l = 0.35 nA kW^{-1} cm^2, was about three times smaller than the CP contribution at the contact region. However, if the light was incident on the bare part of the NW (Fig. 1C), the CP contribution, I_c = −0.1 nA kW^{-1} cm^2, was almost negligible while the LP contribution I_l = 2.2 nA kW^{-1} cm^2 was much stronger, which is in agreement with the previous understanding of LP-dependent photocurrent (*17*, *18*) of semiconductor NWs. The polarization-independent contribution (I_d = 9.6 nA kW^{-1} cm^2) was also larger at the center, because the absorption of light is more than at the NW-metal junction. In another small-diameter device with negligible I_d at the metal-NW junction ($I_d < I_c$), a small contribution from the bulk resulted in a clear reversal in the polarity of the photocurrent with left and right circular polarization (fig. S10).

It is known that Si NWs grown along the ⟨111⟩ and ⟨112⟩ directions terminate on the [110] facets (*21*–*23*), which contain zigzag atomic chains (labeled as the *z* axis in the schematic of Fig. 1A) that are not aligned to the NW growth direction. These states have a different dispersion behavior than the bands in bulk Si and are responsible for its high hole mobility (*24*, *25*). We can understand the origin of the CPGE in our experiments by analyzing the symmetry of this structure. Figure 1A shows the mirror planes (shaded region of the *x*-*y* and *y*-*z* planes) on the [1$\bar{1}$0] surface, which exist in the absence of any electric field. The emergence of chirality can be understood by introducing an electric field along the NW ⟨111⟩ direction, which breaks the mirror symmetries of the *y*-*z* plane (due to the *x* component of the field) and *x*-*y* plane (due to the *z* component of the field). Breaking these mirror symmetries implies that the resulting structure is not identical to its mirror image and hence is a structure with distinct chirality.

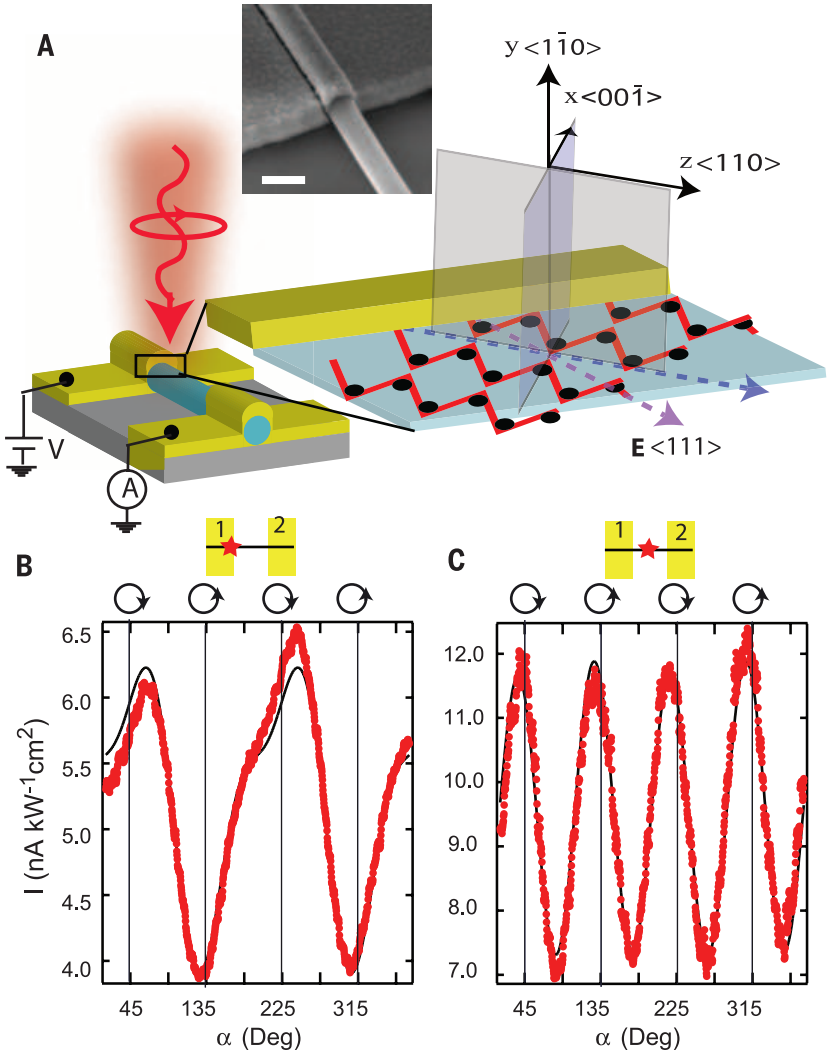


Fig. 1. CP-dependent photocurrent at two different locations along the NW. (**A**) Schematic of the experimental setup (V, voltage source; A, ammeter) along with a schematic of the microscopic view of the metal-semiconductor contact of the Si NW device along with the atomic zigzag chains. The shaded area (*x*-*y* and *y*-*z* planes) represents the mirror planes on the [1$\bar{1}$0]surface. ***E***, direction of the electric field. The inset shows a scanning electron microscopy image of one of the electrodes. Scale bar, 0.2 μm. (**B** and **C**) Photocurrent as a function of the QWP angle is shown at two different positions of laser excitation along the NW (the excitation region is indicated by schematics on the corresponding plots). Photocurrent properties (B) when the laser is incident at the metal-nanowire junction labeled 1 and (C) at the center of the NW are shown. In (B) and (C), the solid lines are the fits to the expression of photocurrent $I(\alpha) = I_c \sin 2\alpha + I_l \cos(4\alpha + \varphi) + I_d$.

In order to study the effect of the Schottky electric field at the metal-semiconductor contact on the CPGE, we measured photocurrents at different applied biases as a function of laser polarization (QWP angle). The CP-dependent contribution to photocurrent (I_c) (Fig. 2A) changes as a function of applied bias. At zero bias, the Schottky field determined by the nature of the metal-semiconductor contact produces photocurrent with a nonzero value of I_c = −66 pA kW^{-1} cm^2, in comparison to a lower value of the LP-dependent coefficient I_l = 37 pA kW^{-1} cm^2, clearly demonstrating the CPGE at the contacts. As the applied bias is modulated, these coefficients change, and the system's response can be tuned (Fig. 2A and table S1) from producing a significant CPGE (at ±248 mV) to almost none (at 80 mV), which clearly shows that the Schottky field at the interface is responsible for the observed effect.

To qualitatively explain the bias-dependent CPGE results, a set of band diagrams is shown in Fig. 2B at different biases. The sign convention of the photocurrent is positive in the direction from junction 1 (J1) to 2 (J2). The NW device can be understood as two back-to-back Schottky contacts; for a *p*-type semiconductor, applying a positive bias at J1 (J2 is grounded) makes J1 reverse-biased and J2 forward-biased. Under dark conditions, the current through the device is limited by the current in the reverse-biased junction, and as a result a negligible dark current is obtained (Fig. 2C). When light is incident at J2, the magnitude of the photocurrent is large if J2 is reverse-biased (J1 is forward-biased), as the photogenerated carriers produce drift current due to the Schottky field at the reverse-biased junction, resulting in an asymmetric current-voltage (*I*-*V*) relation (Fig. 2C, red curve). At zero bias, we observed a CPGE due to the band bending (Fig. 2B). Table S1 summarizes the estimated values of I_c, I_l, and I_d, as a function of applied bias at a laser intensity of 32 W/cm^2. As we increased the bias voltage from zero, J2 was forward-biased, and hence the band bending at J2 decreased, which reduced the CP coefficient; at 80 mV, the CPGE almost vanished (I_c ~ −6 pA kW^{-1} cm^2), and the only remaining contribution was due to I_d and I_l. When the bias voltage was further increased, the electric field direction at J2 was reversed, which reversed the CP contribution, $I_c \rightarrow -I_c$. The sign of the CP-dependent coefficient only depends

on the sign of the electric field because of band bending where light is incident and not on the sign of I_d (the dominant contribution to the photocurrent), as revealed in I_c – V and I_d – V plots (Fig. 2D). Figure 2E shows that the LP-dependent contribution I_l varies little in comparison to I_c as a function of applied bias, which implies that only the CP-dependent contribution is sensitive to applied bias among the two polarization-dependent photocurrent contributions.

To understand the microscopic origin of the symmetry breaking, we adopt the model shown in Fig. 1A. The $[1\bar{1}0]$ plane contains the zigzag chains of Si atoms along the $\langle 110 \rangle$ direction, which makes an angle of ~35° to the growth direction and the Schottky field of the NW. Figure 3A schematically shows a simplified linear chain, where the valence band is formed by p orbitals (p_x, p_y, and p_z) and the conduction band by s orbitals. For small crystal momentum k_z along the z direction, the Schottky electric field is a perturbation that mixes the Bloch states with orthogonal orbital polarizations, $\Psi_1(\boldsymbol{k},\boldsymbol{r}) = (1/\sqrt{N})\Sigma_j \exp(ik_zZ_j)p_x(\boldsymbol{r} - Z_j)$ and $\Psi_2(\boldsymbol{k},\boldsymbol{r}) = (1/\sqrt{N})\Sigma_j \exp(ik_zZ_j)p_z(\boldsymbol{r} - Z_j)$, so that the perturbed states can be written as $\Psi_\pm = \frac{1}{\sqrt{2}}[\Psi_1(\boldsymbol{k},\boldsymbol{r}) \pm i\Psi_2(\boldsymbol{k},\boldsymbol{r})]$ [(*16*), section 5]. An energy-splitting $\sim\pm|k_z\gamma a|$ is obtained between the states Ψ_+ and Ψ_-, where a is the lattice constant and γ is the nearest-neighbor overlap integral between the p_x and p_z orbitals in the presence of a Schottky field, whereas the ± sign depends on the sign of k_z and the Schottky field. Figure 3B shows the splitting of Ψ_+ and Ψ_- energy bands for small values of k_z induced by the Schottky field. Two valence bands $\Psi_\pm$ have unequal orbital population for a nonzero k_z, and as a result, circularly polarized light propagating along the y direction with $\sigma = \pm 1$ can excite electrons only from the initial state with $l_y = \mp 1$, to final state $|S>$ with $l_y = 0$ (Fig. 3B). This generates an asymmetry in the population of momentum distribution of the excited electrons at $\pm k_z$ and hence produces photocurrents, which reverse sign when the chirality of the CP light is changed.

Because a $\langle 111 \rangle$ or $\langle 112 \rangle$ grown NW can also terminate at facets other than the $[1\bar{1}0]$ set of planes (*22*), there is a possibility that the electrically contacted plane is different from $[1\bar{1}0]$ and hence should not produce any CPGE. Statistically, we observed from our measurements that ~40% of NWs did not show any CPGE at the metal-NW contacts but always had the LP-dependent photocurrent. We sometimes observed an asymmetry in the CPGE response (fig. S12) between the two contacts on the same NW, suggesting that contacts may not be symmetric, or the $[1\bar{1}0]$ plane may not run uniformly along the entire length of the NW. This observation is in agreement with the observed stacking faults and other defects typically observed in ~25 to 30% of the as-grown Si NWs, which can disrupt the chains. This statistical analysis implies that the quality of metal-NW contact at the surface of the NWs is important to observe the CPGE, which is further corroborated by the observation that the CPGE phenomenon decreases as the device ages, probably due to the diffusion of the metal or other impurities that can disrupt the chains.

Although our model explains the bias-dependent results, we performed more experiments, such as the power and energy dependence of the photocurrent, for further verification. Figure 3C shows a linear dependence of the CP contribution (I_c) with light intensity, which is required for the CPGE as $I_c \propto i(\mathcal{E}_x\mathcal{E}_z^* - \mathcal{E}_z\mathcal{E}_x^*)$, where $\mathcal{E}_x$ and $\mathcal{E}_z$ are complex electric field components of light. The

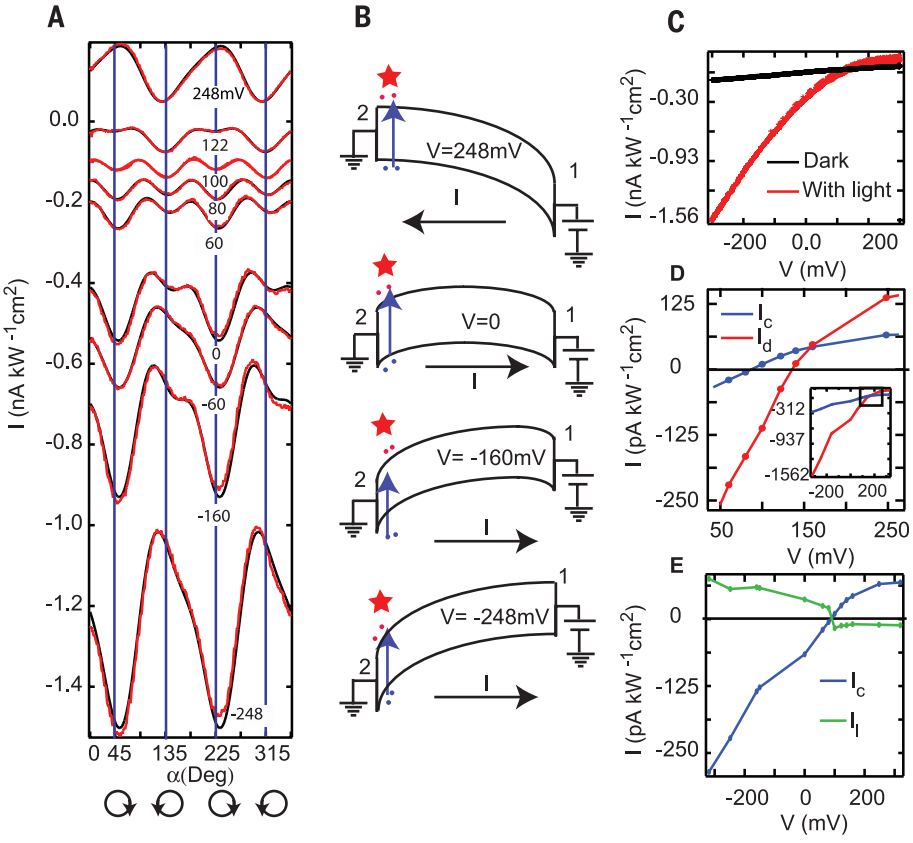


Fig. 2. Effect of bias-tunable Schottky field on CP-dependent photocurrents. (**A**) Photocurrent as a function of QWP angle (α) at different applied biases. (**B**) Schematic band diagrams at four representative bias voltages for a *p*-type NW. (**C**) *I*-*V* curves of the device in the dark (black curve) and upon light excitation at the reverse-biased junction (red curve). (**D**) Zoomed-in view (close to zero current) of the I_c and I_d (obtained from fits for $I(\alpha)$, table S1) as a function of applied bias (V). (**E**) Plot of I_c and I_l as a function of applied bias.

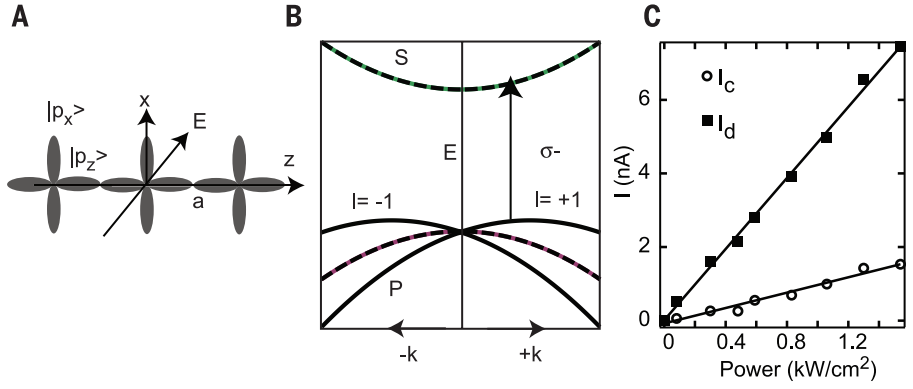


Fig. 3. Orbital-based linear chain model for the CPGE in Si nanowires. (**A**) Schematic of a linear chain of atoms along the z direction with the valence band formed by p orbitals. Light is normally incident along the y direction. The direction of the electric field (***E***) due to Schottky contact and the growth direction of the NW (along $\langle 111 \rangle$) lie in the x-z plane. (**B**) Plots of calculated energy band dispersion using the linear chain model, showing the splitting of valence bands due to the mixing of $|p_x\rangle$ and $|p_z\rangle$ orbitals, in the presence of the Schottky field [(*16*), section 5]. (**C**) The power dependence of the CPGE (open circles) as well as the polarization-independent background (solid squares) shows linear dependence with laser power.

polarization-independent contribution (I_d) is also linear with the light intensity, and hence we used this quantity to normalize the I_c to study the wavelength dependence of the CPGE.

To verify the hypothesis that [110] surface states contribute to the observed CPGE, we performed laser wavelength–dependent photocurrent measurements in the energy range from 500 to 800 nm (Fig. 4A) and plotted a normalized quantity, I/I_d, as a function of the QWP angle. It is expected that the absorption in the bulk Si NW should mostly affect the background and LP-dependent photocurrent, whereas the surface states, because of their particular dispersions, may have a different response. Previous studies via angle-resolved photoemission spectroscopy (*24*, *25*) and scanning tunneling microscopy (*26*) measurements have demonstrated that the Si [110] plane contains many surface states with varying energy gaps. An identification of these with the states at the NW surface is only approximate, because the surface states are sensitive owing to surface reconstruction. Nevertheless, the energy difference between some of the reported states matches (within the thermal energy) the excitation energy of the laser at which a strong signal is recorded. For example, the energy difference of surface states *C3* (2.6 eV) and *S3* (0.75 eV) (*24*) matches the laser excitation at 680 nm, where we see a larger I_c/I_d as compared to 500 and 710 nm. We also measured the response at lower excitation energies and observed that the CPGE response is stronger at 800 nm in addition to 680 nm (Fig. 4A), which matches the energy difference between the *C3* (2.6 eV) and *S4* (1 eV) surface states (*24*). The quantity I_c/I_d that characterizes the relative strength of the CPGE over the background contribution as a function of excitation wavelength (Fig. 4A) changes from a maximum value of ~0.19 observed at 800 nm to 0.02 at 500 nm (the minimum obtained). We studied the polarization-dependent photocurrents at 77 K and observed an overall decrease in the photocurrent, which is expected because of the reduction of phonon-induced absorption in bulk Si as well as the reduction of thermally generated carriers; however, the overall magnitude of normalized I_c at 77 K was similar to the room temperature results (Fig. 4, B and C). More importantly, we observed a blue shift of the excitation energy corresponding to a maximum in I_c/I_d (Fig. 4, B and C), which suggests that the energy bandgaps of the surface states increase at lower temperature. The above observations imply that our experiments support the idea of extended one-dimensional states at the surface of Si NWs grown along the ⟨111⟩ or ⟨112⟩ direction, responsible for the observed CPGE and not dependent on metal antenna, plasmonic, or hot electron effects (*19*, *20*).

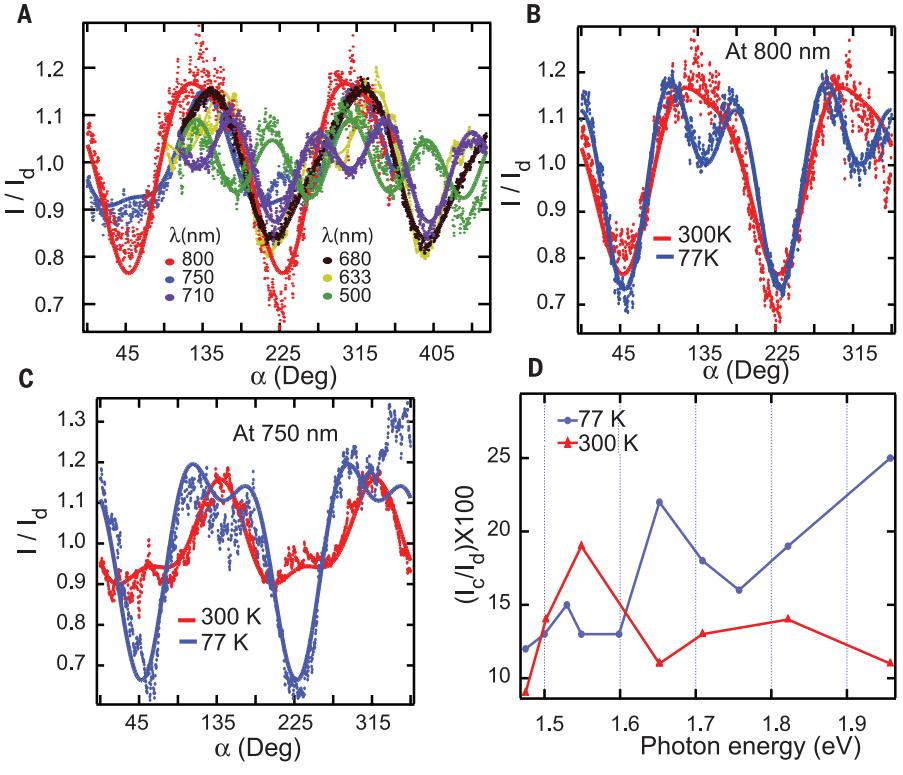


Fig. 4. Wavelength and temperature dependence of the CPGE. (**A**) Normalized photocurrents as a function of QWP angle α for different excitation wavelengths in the range of 500 to 800 nm at room temperature along with fits (solid lines). Comparisons between polarization-dependent normalized photocurrents obtained at room temperature and at 77 K for (**B**) 800-nm and (**C**) 750-nm laser excitation are shown. (**D**) Trend of I_c per unit of I_d given in percentage ($\frac{I_c}{I_d} \times 100$) as a function of laser excitation wavelength at room temperature and at 77 K.

Si NWs support the CPGE where the propagation direction of the current can be controlled by the sense of circular polarization of light, which is also tunable with an external bias. The CPGE traditionally occurs in gyrotropic optical media and hence does not occur in bulk Si because of its diamond structure. In contrast, we showed that the CPGE in Si NWs is a purely orbital effect, and it arises from the geometrical effect in the NW that reduces the bulk mirror symmetries. Because the effect can be engineered by a combination of shape, crystal anisotropy, and applied fields, many applications, as well as the ability to encode more information in a device by using the orbital degrees of freedom, are possible.

REFERENCES AND NOTES

1. B. I. Sturman, V. M. Fridkin, *The Photovoltaic and Photorefractive Effects in Non-Centrosymmetric Materials* (Gordon and Breach Science Publishers, New York, 1992).
2. S. D. Ganichev *et al.*, *Nature* **417**, 153–156 (2002).
3. S. A. Tarasenko, E. L. Ivchenko, *JETP Lett.* **81**, 231–235 (2005).
4. E. Ivchenko, S. Tarasenko, *Semicond. Sci. Technol.* **23**, 114007 (2008).
5. S. Ganichev, W. Prettl, *J. Phys. Condens. Matter* **15**, R935–R983 (2003).
6. S. Tarasenko, *JETP Lett.* **85**, 182–186 (2007).
7. J. Karch *et al.*, *J. Phys. Condens. Matter* **22**, 355307 (2010).
8. J. Karch *et al.*, *Phys. Rev. B* **83**, 121312(R) (2011).
9. E. L. Ivchenko, B. Spivak, *Phys. Rev. B* **66**, 155404 (2002).
10. P. De Padova *et al.*, *Nano Lett.* **8**, 271–275 (2008).
11. B. Aufray *et al.*, *Appl. Phys. Lett.* **96**, 183102 (2010).
12. A. Kara *et al.*, *Surf. Sci. Rep.* **67**, 1–18 (2012).
13. T. Satô, Y. Takeishi, H. Hara, Y. Okamoto, *Phys. Rev. B* **4**, 1950–1960 (1971).
14. A. Ciucivara, B. R. Sahu, S. Joshi, S. K. Banerjee, L. Kleinman, *Phys. Rev. B* **75**, 113309 (2007).
15. V. W. C. Min Yang *et al.*, *IEEE Trans. Electron. Dev.* **53**, 965–978 (2006).
16. Materials and methods are available as supporting materials on *Science* Online.
17. J. Wang, M. S. Gudiksen, X. Duan, Y. Cui, C. M. Lieber, *Science* **293**, 1455–1457 (2001).
18. Y. Ahn, J. Dunning, J. Park, *Nano Lett.* **5**, 1367–1370 (2005).
19. P. Fan, K. C. Y. Huang, L. Cao, M. L. Brongersma, *Nano Lett.* **13**, 392–396 (2013).
20. M. W. Knight, H. Sobhani, P. Nordlander, N. J. Halas, *Science* **332**, 702–704 (2011).
21. R. Rurali, *Rev. Mod. Phys.* **82**, 427–449 (2010).
22. R. Q. Zhang *et al.*, *J. Chem. Phys.* **123**, 144703 (2005).
23. J. F. Justo, R. D. Menezes, L. V. C. Assali, *Phys. Rev. B* **75**, 045303 (2007).
24. K. Sakamoto *et al.*, *Phys. Rev. B* **79**, 045304 (2009).
25. A. Crincenti, B. Nesterenko, P. Perfetti, G. LeLay, C. Sebenne, *J. Vac. Sci. Technol. A* **14**, 2448–2453 (1996).
26. M. Setvín *et al.*, *Phys. Rev. B* **84**, 115317 (2011).

ACKNOWLEDGMENTS

This work was supported by the U.S. Army Research Office (grants W911NF-09-1-0477 and W911NF-11-1-0024) and the seed project support from the Laboratory for Research on the Structure of Matter, NSF Materials Research Science and Engineering Center grant DMR-1120901. E.J.M. is supported by the U.S. Department of Energy–Basic Energy Sciences under grant DE FG02 84ER45118. Si NWs were provided by B. Tian (Chicago) and CINT, a U.S. Department of Energy, Office of Basic Energy Sciences User Facility at Los Alamos National Laboratory (contract DE-AC52-06NA25396) and Sandia National Laboratories (contract DE-AC04-94AL85000). The data described in the paper are archived by the Agarwal Group at the University of Pennsylvania.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/726/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1

21 May 2015; accepted 10 July 2015
Published online 23 July 2015
10.1126/science.aac6275

GLIAL CELL SIGNALING

Circuit-specific signaling in astrocyte-neuron networks in basal ganglia pathways

R. Martín,[1]* R. Bajo-Grañeras,[1]* R. Moratalla,[1,2] G. Perea,[1] A. Araque[3]†

Astrocytes are important regulatory elements in brain function. They respond to neurotransmitters and release gliotransmitters that modulate synaptic transmission. However, the cell- and synapse-specificity of the functional relationship between astrocytes and neurons in certain brain circuits remains unknown. In the dorsal striatum, which mainly comprises two intermingled subtypes (striatonigral and striatopallidal) of medium spiny neurons (MSNs) and synapses belonging to two neural circuits (the direct and indirect pathways of the basal ganglia), subpopulations of astrocytes selectively responded to specific MSN subtype activity. These subpopulations of astrocytes released glutamate that selectively activated *N*-methyl-D-aspartate receptors in homotypic, but not heterotypic, MSNs. Likewise, astrocyte subpopulations selectively regulated homotypic synapses through metabotropic glutamate receptor activation. Therefore, bidirectional astrocyte-neuron signaling selectively occurs between specific subpopulations of astrocytes, neurons, and synapses.

Astrocytes respond to synaptically released neurotransmitters and release gliotransmitters that regulate synaptic transmission [(*1*–*6*), but see (*7*–*8*)]. However, the question of whether this functional relationship is cell- and synapse-specific remains unexplored. To address this issue, we studied the dorsal striatum, which presents unique structural and functional properties, because it mainly comprises two subtypes of intermingled but molecularly identifiable neurons [striatonigral and striatopallidal medium spiny neurons (MSNs)] and synapses belonging to two distinct neural circuits (the basal ganglia's direct and indirect pathways). Hence, individual MSNs can be selectively stimulated while analyzing the astrocytic activity and the consequent neuromodulation of particular neurons and synapses.

We first performed paired recordings of MSNs, stimulating one neuron by depolarization and monitoring corticostriatal excitatory postsynaptic currents (EPSCs) in that neuron (homoneuronal synapses) as well as in the adjacent neuron (heteroneuronal synapses) (Fig. 1A and supplementary materials). The stimulation of single MSNs by depolarizing pulses or action potential trains (fig. S1), which trigger endocannabinoid (eCB) release (*9*–*12*), induced a transient depression (76.5 ± 2.9% relative to the control EPSC amplitude, $P < 0.001$) in 16 out of 33 homoneuronal synapses (48.5%) (Fig. 1, B and C). This synaptic depression was abolished by the cannabinoid receptor type 1 (CB1R) antagonist AM251 (2 μM, $n = 10$ MSN pairs), was absent in slices from CB1R-deficient ($CB1R^{-/-}$) mice ($n = 24$ MSN pairs) (Fig. 1C), and was associated with an increase in the paired-pulse ratio (PPR; $n = 16$ MSN pairs) (fig. S1G). This indicates that the depression corresponded to a depolarization-induced suppression of excitation (DSE) mediated by the activation of presynaptic CB1Rs (*13*).

In contrast, in simultaneously recorded heteroneuronal synapses, MSN stimulation induced a transient potentiation (138.5 ± 7.7% relative to the control EPSC amplitude, $P < 0.001$) in 7 out of 16 synapses (43.8%) (Fig. 1, B and C) that concomitantly evoked DSE in homoneuronal synapses. No changes were detected in the rest of the heteroneuronal synapses (fig. S2). This potentiation was abolished by AM251 ($n = 7$ MSN pairs), was absent in $CB1R^{-/-}$ mice ($n = 24$) (Fig. 1C), and was accompanied by a PPR decrease ($n = 7$ MSN pairs) (fig. S1G), indicating that it was also mediated by CB1R activation. Because the eCB-mediated DSE has been well characterized in MSNs (*9*, *12*), we focused our study on the heteroneuronal synaptic responses.

The two types of MSNs are not spatially segregated, but they express different dopamine receptors. Their projections form the origin of the direct and indirect pathways of the basal ganglia motor circuit. Direct-pathway striatonigral MSNs express D1 receptors, and indirect-pathway striatopallidal MSNs express D2 receptors (termed D1 and D2 MSNs, respectively) (*14*, *15*). We investigated whether eCB-mediated heteroneuronal synaptic potentiation was specific of the neuronal subtypes, using Drd1a-tdTomato and Drd2-EGFP bacterial artificial chromosome (BAC) transgenic mice that respectively allowed the identification of the D1 and D2 MSNs (*16*, *17*) (Fig. 1D). We recorded homotypic (D1-D1 or D2-D2) and heterotypic (D1-D2) pairs of MSNs, stimulating one neuron to evoke eCB release and analyzing synaptic transmission in the other neuron (Fig. 1E). Stimulation of either D1 or D2 MSNs induced a heteroneuronal potentiation selectively in homotypic pairs ($n = 10$ out of 11 D1-D1 pairs and 11 out of 12 D2-D2 pairs), but no synaptic changes were observed in heterotypic MSNs ($n = 13$ D1-D2 and 12 D2-D1 pairs in which D1 and D2 MSNs were stimulated, respectively) (Fig. 1F). This heteroneuronal potentiation between homotypic neurons was abolished by AM251 ($n = 10$ D1-D1 and 11 D2-D2 pairs) and absent in mice produced by crossing Drd1a-tdTomato and $CB1R^{-/-}$ mice (D1-Tom-$CB1R^{-/-}$; $n = 13$ D1-D1 and 12 nonfluorescence and putatively D2-D2 pairs) (Fig. 1G). The different synaptic regulation of homotypic and heterotypic neurons was not due to differences in eCB release, because only neuronal pairs that showed homoneuronal DSE were considered (fig. S3). Taken together, these results indicate cell-specific signaling between MSN subtypes belonging to the direct or indirect basal ganglia pathways.

The synaptic potentiation of homotypic MSNs was abolished by antagonists of the group I metabotropic glutamate receptors ($mGluR_{5/1}$) MPEP (50 μM) and LY367385 (100 μM), suggesting the participation of glutamate ($n = 10$ D1-D1 and 11 D2-D2 pairs) (Fig. 1G). Consistent with a presynaptic mechanism suggested by PPR changes, in paired recordings of homotypic MSNs in which one neuron was loaded with guanosine 5′-*O*-(2′-thiodiphosphate) (GDP-β-S; 2 mM) to prevent postsynaptic mGluR-mediated signaling (confirmed by the absence of calcium elevations in response to application of the $mGluR_{5/1}$ agonist DHPG) (Fig. 2, A to C), neuronal stimulation similarly evoked heteroneuronal potentiation, enhanced miniature EPSC (mEPSC) frequency but not amplitude, and decreased PPR in control and GDP-β-S–loaded neurons (Fig. 2, D to F). Moreover, local application of DHPG mimicked these responses (Fig. 2, D to F), further supporting an mGluR-dependent presynaptic mechanism.

The neuronal release of eCBs can activate astrocytic CB1Rs and stimulate glutamate release in the hippocampus and cortex (*18*–*20*). We thus tested whether astrocytes responded to eCBs released from MSNs by monitoring astrocyte Ca^{2+} levels in response to neuronal stimulation (Fig. 2, G and H, and fig. S4). MSN depolarization elevated Ca^{2+} in 71 out of 153 astrocytes (46.4% of astrocytes, 16 slices), increasing both the Ca^{2+}-spike probability and the oscillation frequency (Fig. 2, G to H). This astrocyte Ca^{2+} signal was abolished by AM251 and was absent in $CB1R^{-/-}$ and inositol 1,4,5-trisphosphate receptor type 2–deficient ($IP_3R2^{-/-}$) mice, in which heterotrimeric guanine nucleotide-binding protein (G protein)–mediated Ca^{2+} elevations are selectively impaired in astrocytes (*18*, *21*–*23*). Therefore, eCBs released from MSNs elevated astrocytic Ca^{2+} through CB1R activation (Fig. 2H).

We then tested whether the heteroneuronal potentiation required the astrocyte Ca^{2+} signal. This potentiation was absent in $IP_3R2^{-/-}$

[1]Instituto Cajal, Consejo Superior de Investigaciones Científicas, 28002 Madrid, Spain. [2]Centro de Investigación Biomédica en Red Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, 28029 Madrid, Spain. [3]Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA.
*These authors contributed equally to this work.
†Corresponding author. E-mail: araque@umn.edu

mice (n = 28) (Fig. 2I); however, DSE recorded in homoneuronal synapses was unaffected (fig. S5) (*18*), indicating that the eCB-release machinery and neuronal CB1R activation was preserved in these animals. Conversely, the selective increase of Ca^{2+} levels by ultraviolet (UV)–flash photolysis of the Ca^{2+} cage *o*-nitrophenyl-EGTA (NP-EGTA), loaded in whole-cell recorded astrocytes, potentiated the synaptic transmission.

We next monitored *N*-methyl-D-aspartate receptor (NMDAR)–dependent neuronal slow inward currents (SICs) (*24*–*26*), which are a biological assay to detect astrocytic glutamate release (*1*). In MSN-pair recordings, stimulation of one neuron elevated astrocyte Ca^{2+} levels and increased the frequency of SICs in adjacent MSNs (Fig. 2; G, J, and K). SICs were abolished by the NMDAR antagonist AP5 (50 μM), without affecting the astrocyte Ca^{2+} signals (n = 11 MSN pairs), and were unaffected by tetrodotoxin (1 μM, n = 8 MSN pairs) (Fig. 2L). These results indicate that striatal astrocytes may release glutamate upon elevating their Ca^{2+} levels (fig. S6). Local application of the CB1R agonist WIN55,212-2 (WIN; 100 μM) mimicked the neuron-evoked increase of both astrocyte Ca^{2+} levels and SIC frequency (n = 15 MSNs) (fig. S7). Neuronal- and WIN-evoked Ca^{2+} elevations and SICs were abolished by AM251 and absent in $CB1R^{-/-}$ and $IP_3R2^{-/-}$ mice (Fig. 2H and fig. S7). Hence, eCBs released from MSNs activate CB1Rs in astrocytes, which elevate their intracellular Ca^{2+} and stimulate the release of glutamate that activates neuronal NMDARs. In aggregate, these findings indicate that MSNs signal to astrocytes through eCBs, and, in turn, astrocytes signal to MSNs through glutamate.

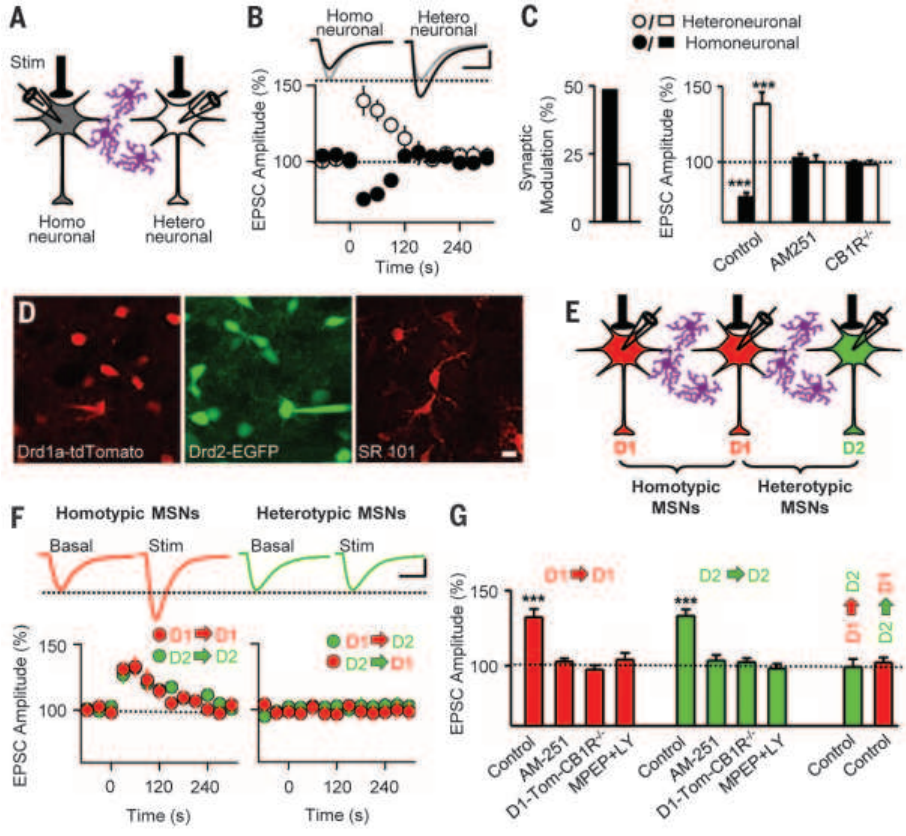


Fig. 1. Differential modulation of synaptic transmission by eCBs in the dorsal striatum. (**A**) Scheme showing depolarization of one MSN while EPSCs were monitored in that neuron (homoneuronal) and in the adjacent neuron (heteroneuronal). Astrocytes are shown in purple. (**B**) Representative EPSCs (top) before (gray) and after (black) neuronal stimulation, with relative responses (bottom) in homoneuronal (black circles) and heteroneuronal (white circles) synapses. On the *x* axis, 0 s corresponds to the onset of the neuronal stimulation, as in all other figures. The scale bars are 25 pA (vertical) and 10 ms (horizontal). (**C**) The percentage of homoneuronal and heteroneuronal synapses showing modulation (left) and relative EPSC amplitude in control, AM251, and $CB1R^{-/-}$ mice (each bar: $n \geq 7$ synapses, $n \geq 7$ slices) (right). (**D**) Fluorescence images of MSNs from Drd1a-tdTomato and Drd2-EGFP BAC-transgenic mice and sulforhodamine 101 (SR101)–stained astrocytes. The scale bar is 10 μm. (**E**) Scheme of our experimental paradigm. (**F**) Averaged EPSCs (top) before (basal) and after D1- or D2-MSN stimulation (stim) in homotypic and heterotypic MSNs, with EPSC amplitudes (bottom). The scale bars are 25 pA (vertical) and 15 ms (horizontal). (**G**) Relative EPSC amplitude changes in control, AM251, D1-Tom-$CB1R^{-/-}$ mice, and MPEP+LY367385 (MPEP+LY) (each bar: $n \geq 10$ MSNs, $n \geq 10$ slices). ***$P < 0.001$; Student's t test. Data are expressed as mean ± SEM.

We then studied whether the reciprocal signaling between astrocytes and MSNs was cell type–specific and pathway-specific. First, we investigated whether subpopulations of astrocytes were functionally associated with MSN subtypes by monitoring astrocytic Ca^{2+} levels and recording pairs of identified homotypic or heterotypic MSNs. The number of responding astrocytes, as well as the increase in the Ca^{2+}-spike probability and oscillation frequency, were similar upon stimulation of D1 and D2 MSNs (Fig. 3, A to D). However, whereas 55.4 ± 7.0% (n = 117 astrocytes, 14 slices) and 46.1 ± 4.8% (n = 129 astrocytes, 12 slices) of the astrocytes responded to depolarization of homotypic D1-D1 or D2-D2 MSNs, respectively, only 10.7 ± 3.2% of the astrocytes (n = 118 astrocytes, 14 slices) responded to stimulation of both types of D1-D2 MSNs in heterotypic pairs (Fig. 3, E to G). Subsets of astrocytes thus selectively responded to the activity of either D1 or D2 MSNs, which suggests the existence of specific neuron-to-astrocyte signaling between specific MSN subtypes and particular striatal astrocytes.

We then tested whether the astrocyte-to-neuron signaling shown in Fig. 2 was restricted to particular neuronal subtypes by analyzing SICs in identified MSN pairs. In homotypic pairs, stimulation of one neuron increased SIC frequency in adjacent MSNs (n = 12 D1-D1 and 15 D2-D2 pairs). In contrast, in heterotypic pairs, MSN stimulation failed to increase the SIC frequency (n = 17 D1-D2 and 13 D2-D1 pairs, where D1 or D2 MSNs were stimulated and SICs were recorded from D2 or D1 MSNs, respectively) (Fig. 3, H and I). This indicates that the nonsynaptic astrocyte-mediated communication between MSNs was specific of homotypic cell subtypes and restricted to particular striatal pathways. These results suggest the existence of functional astro-neuronal networks defined by the presence of selective astrocyte-neuron bidirectional signaling between specific cell subtypes.

We therefore asked whether subpopulations of astrocytes regulated specific subsets of synapses belonging to the direct and indirect basal ganglia pathways. We performed triple whole-cell recordings of sets of two heterotypic MSNs with a single astrocyte. We first identified astrocytes that responded to the depolarization of either D1 or D2 MSNs. Then, an identified astrocyte was loaded with NP-EGTA (5 mM) to be stimulated by UV-flash photolysis. Finally, another pair of heterotypic neurons was recorded to monitor the synaptic responses (Fig. 4, A to C). Uncaging of Ca^{2+} in astrocytes that responded to D1-MSN stimulation evoked a transient synaptic potentiation exclusively in D1 MSNs (n = 7 out of 9 triple recordings) without affecting neurotransmission in D2 MSNs (n = 9 out of 9 triple recordings). Likewise, selective UV-flash stimulation of D2-responding astrocytes selectively enhanced synaptic transmission in D2 (n = 8 out of 9) but not in D1 MSNs (n = 9 out of 9) (Fig. 4, D and E). These results indicate that astrocyte-mediated synaptic regulation is restricted to signaling between subsets of astrocytes and particular synapses.

Although some reports have questioned the physiological significance of the astrocyte Ca^{2+} signal and its neuromodulatory consequences (*7*, *8*), our findings reveal that this signal can be triggered by endogenous stimuli of neuronal origin and that its effects on synaptic function are delicately regulated.

Astrocytes and neurons in the dorsal striatum selectively interact in a cell- and synapse-specific manner, and striatal astrocytes display a functional

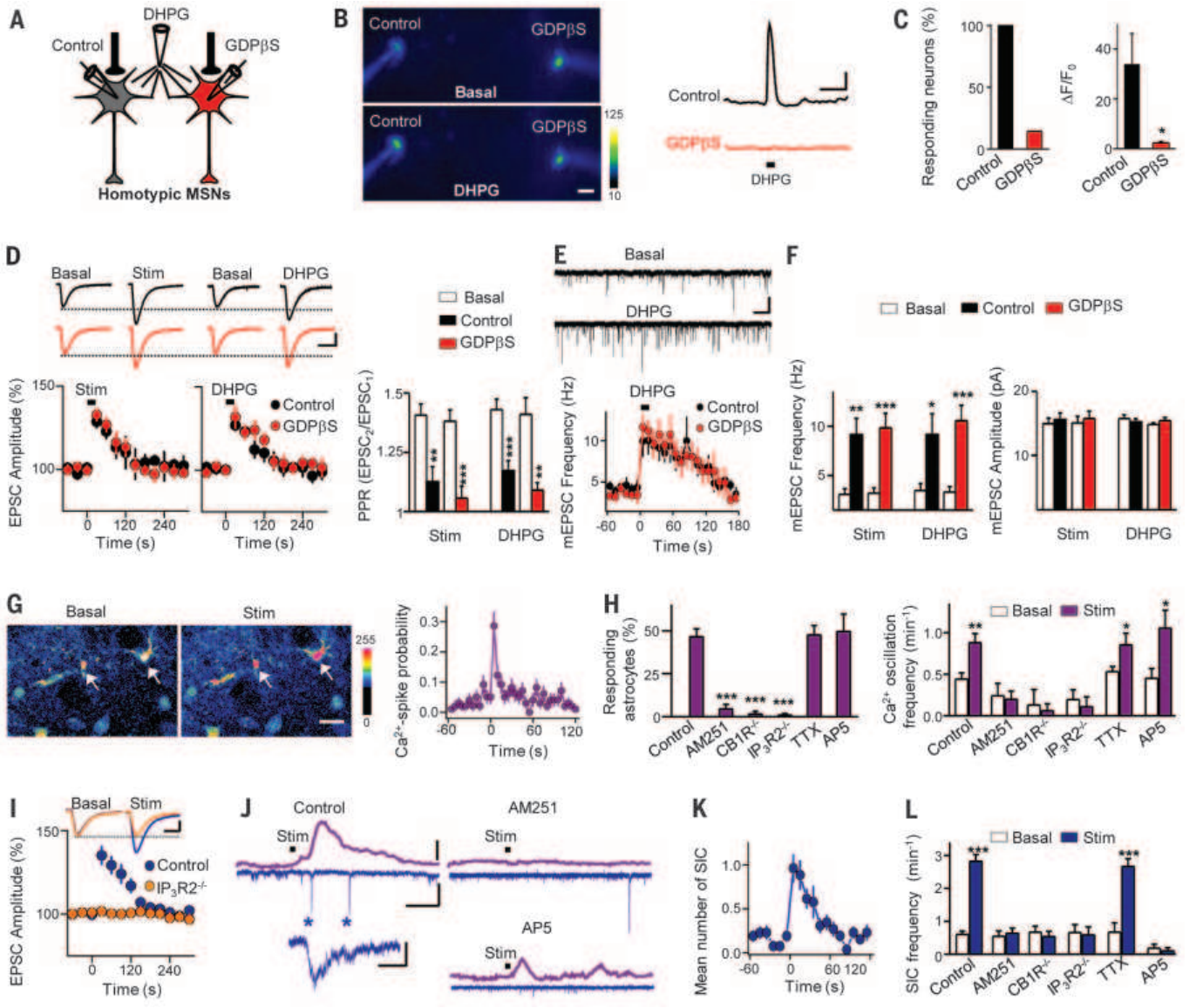


Fig. 2. Bidirectional astrocyte-neuron signaling regulates MSN excitability and synaptic transmission. (**A**) Scheme showing homotypic MSN pair recordings. One neuron was loaded with GDP-β-S. (**B**) Pseudocolor images of homotypic MSNs loaded with fluo-4 (control) and fluo-4 and GDP-β-S before (basal) and after local application of DHPG (scale bar, 20 μm). Colors correspond to fluorescence levels. MSN Ca^{2+} responses are shown to the right [scale bars, 20% (vertical) and 20 s (horizontal)], with the application of DHPG indicated by the black bar. (**C**) Responding neurons and increases in fluorescence evoked by DHPG in control and GDP-β-S–loaded MSNs (n = 7 MSN pairs). (**D**) EPSCs before and after homotypic MSN stimulation and DHPG application (top). The bottom left panel shows relative EPSC amplitudes over time. The bottom right panel shows PPR before and after stimulation and DHPG application in control (n = 8 stim and 8 DHPG) and GDP-β-S–loaded (n = 10 stim and 11 DHPG) neurons. The scale bars are 25 pA (vertical) and 15 ms (horizontal). (**E**) MSN whole-cell currents showing mEPSCs before and after DHPG application (top), with mean mEPSC frequency over time before and after DHPG application (black bar) in control (n = 8) and GDP-β-S–loaded (n = 8) neurons (bottom). The scale bars are 25 pA (vertical) and 5 s (horizontal). (**F**) mEPSC frequency and amplitude 1 min before and after stimulation and DHPG application in control and GDP-β-S–loaded neurons (each bar: n ≥ 8 MSN pairs). (**G**) Pseudocolor images of fluo-4–filled astrocytes (indicated by arrows) before and after MSN stimulation, with astrocyte Ca^{2+}-spike probability shown to the right. The scale bar is 10 μm. (**H**) The proportion of astrocytes that responded to neuronal stimulation (left) and Ca^{2+} oscillation frequency (right) before and after stimulation in control and various experimental configurations (each bar: n ≥ 71 astrocytes, n ≥ 9 slices). (**I**) EPSCs before and after stimulation in control (n = 8 MSN pairs, 8 slices) and $IP_3R2^{-/-}$ (n = 28 MSN pairs, 28 slices) mice, with relative EPSC amplitude changes. The scale bars are 25 pA (vertical) and 15 ms (horizontal). (**J**) Representative astrocyte Ca^{2+} levels (purple; scale bar, 20%) and MSN whole-cell currents [blue; scale bars, 25 pA (vertical) and 30 s (horizontal)] in control, AM251, and AP5. Black bars indicate stimulation. Asterisks indicate SICs. The insert shows an expanded SIC [scale bars, 25 pA (vertical) and 250 ms (horizontal)]. (**K**) Mean number of SICs over time. (**L**) Mean SIC frequency before and after stimulation in different experimental configurations (each bar: n ≥ 8 MSN pairs). *P < 0.05, **P < 0.01, ***P < 0.001; Student´s t test. Data are expressed as mean ± SEM.

heterogeneity based on selective signaling with particular neuron subtypes and synapses belonging to the basal ganglia's direct and indirect pathways (Fig. 4F). The activities of both pathways are fundamental in motor control (*27*), and their imbalances underlie motor deficits in Parkinson's (*10*, *11*, *28*, *29*) and Huntington's diseases (*30*). Selective regulation of specific synapses by astrocytes indicates that they may be involved in the

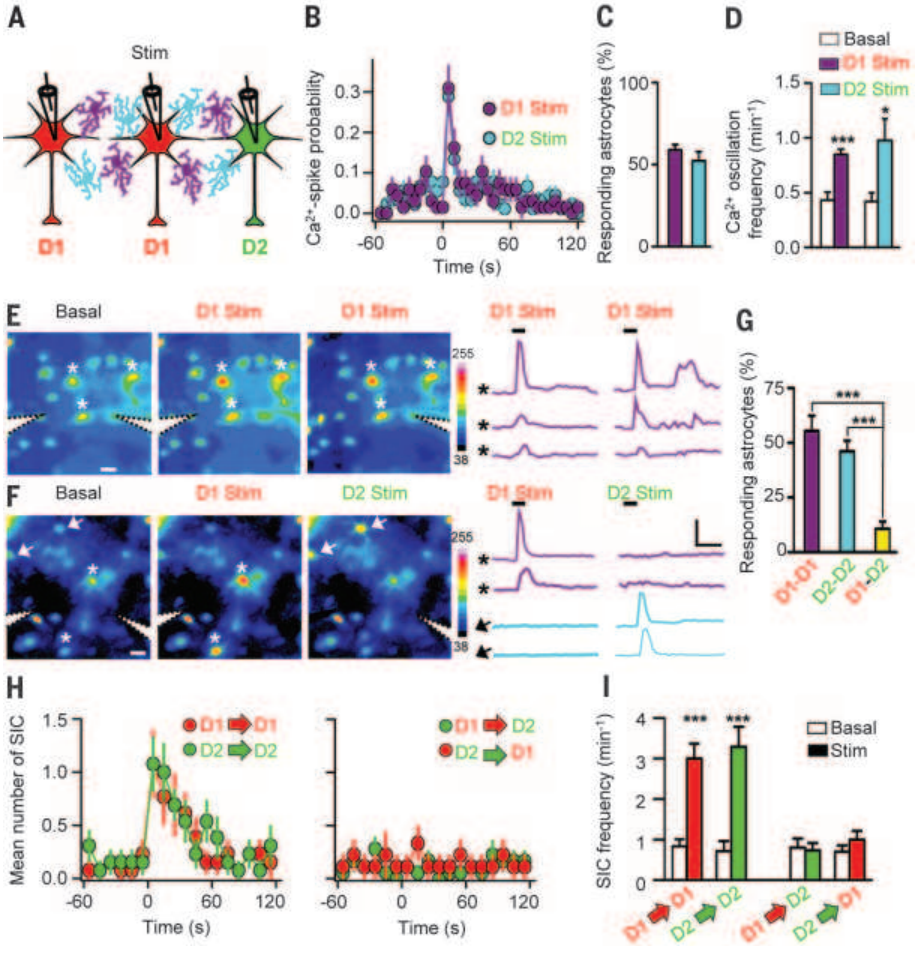


Fig. 3. Astrocyte-neuron signaling is cell type–specific. (**A**) Scheme showing stimulation of one MSN and synaptic currents recorded in homotypic or heterotypic MSNs while monitoring astrocyte Ca^{2+} levels (purple, D1-MSN-responding astrocytes; blue, D2-MSN–responding astrocytes). (**B**) Ca^{2+}-spike probability for astrocytes, (**C**) proportion of responding astrocytes, and (**D**) Ca^{2+} oscillation frequency for astrocytes before (basal) and after stimulation of D1 or D2 MSNs (each bar: $n \geq 235$ astrocytes, $n \geq 26$ slices; $^{*}P < 0.05$, $^{***}P < 0.001$; Student´s *t* test). (**E**) Pseudocolor images (left) of astrocytes (indicated by asterisks; scale bar, 10 μm) and corresponding Ca^{2+} responses (right) before and after consecutive stimulation of a homotypic D1 MSN pair. On the left, colors correspond to fluorescence levels. On the right, horizontal bars indicate neuronal stimulation [vertical and horizontal scales follow (F)]. (**F**) As in (E), but stimulating a heterotypic MSN pair. Asterisks and arrows indicate astrocytes responding to D1- and D2-MSN stimulation, respectively. The scale bars are 20% (vertical) and 10 s (horizontal). (**G**) Percentage of astrocytes responding to consecutive stimulation of homotypic and heterotypic MSNs (each bar: $n \geq 117$ astrocytes, $n \geq 12$ slices; $^{***}P < 0.001$; Fisher´s test). (**H**) Mean number of SICs before and after D1- or D2-MSN stimulation, recorded in homotypic and heterotypic adjacent MSNs. (**I**) Mean SIC frequency recorded in D1 and D2 MSNs before and after neuronal stimulation of homotypic and heterotypic MSNs (each bar: $n \geq 12$ MSN pairs from $n \geq 12$ slices; $^{***}P < 0.001$; Student´s *t* test). Data are expressed as mean ± SEM.

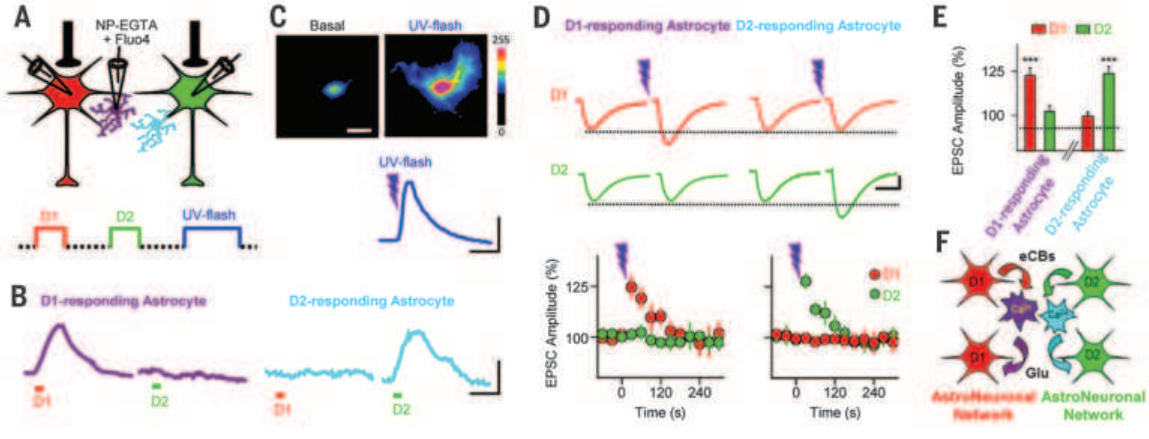


Fig. 4. Synapse-specific regulation of striatal transmission by selective stimulation of particular astrocytes. (**A**) Scheme depicting triple recordings of two heterotypic MSNs and one NP-EGTA– and fluo-4–filled astrocyte (top); the D1- and D2-MSN stimulation protocol for identifying responding astrocytes (consecutive depolarization of D1 and D2 MSNs) and for UV-flash astrocyte stimulation (bottom). (**B**) Intracellular Ca^{2+} levels of two single astrocytes selectively responding to D1- or D2-MSN stimulation. The scale bars are 10% (vertical) and 20 s (horizontal). (**C**) Pseudocolor images of a fluo-4–filled astrocyte before (basal) and after UV-flash stimulation (top; scale bar, 10 μm) and corresponding Ca^{2+} levels [bottom; scale bars, 10% (vertical) and 20 s (horizontal)]. Colors correspond to fluorescence levels. (**D**) Averaged EPSCs [top; scale bars, 25 pA (vertical) and 15 ms (horizontal)], EPSC amplitude changes (bottom), and (**E**) relative EPSC amplitude changes recorded from D1 and D2 MSNs before and after UV-flash stimulation of D1- or D2-responding astrocytes ($n = 9$ and 9; $^{*}P < 0.05$, $^{***}P < 0.001$; Student´s *t* test). Data are expressed as mean ± SEM. (**F**) Scheme representing striatal astro-neuronal networks. Stimulation of MSNs (D1 or D2) promotes eCB release that increases Ca^{2+} in a specific subpopulation of astrocytes, which then release glutamate that modulates excitability and synaptic transmission selectively in homotypic neurons through activation of NMDARs and group I mGluRs, respectively.

coordinated activity of these networks in the striatal function, and, therefore, they may participate in its dysfunction in brain disorders. Our results demonstrate the existence of functional astro-neuronal networks that comprise subpopulations of astrocytes, neurons, and synapses belonging to specific brain circuits, which may differentially control specific circuit activity through selective signaling between particular astrocytes and neurons.

REFERENCES AND NOTES

1. A. Araque *et al.*, *Neuron* **81**, 728–739 (2014).
2. C. Eroglu, B. A. Barres, *Nature* **468**, 223–231 (2010).
3. M. M. Halassa, P. G. Haydon, *Annu. Rev. Physiol.* **72**, 335–355 (2010).
4. G. Perea, M. Navarrete, A. Araque, *Trends Neurosci.* **32**, 421–431 (2009).
5. A. Volterra, J. Meldolesi, *Nat. Rev. Neurosci.* **6**, 626–640 (2005).
6. G. Perea, A. Araque, *Science* **317**, 1083–1086 (2007).
7. C. Agulhon, T. A. Fiacco, K. D. McCarthy, *Science* **327**, 1250–1254 (2010).
8. T. A. Fiacco *et al.*, *Neuron* **54**, 611–626 (2007).
9. B. C. Shonesy *et al.*, *Nat. Neurosci.* **16**, 456–463 (2013).
10. A. C. Kreitzer, R. C. Malenka, *Nature* **445**, 643–647 (2007).
11. W. Shen, M. Flajolet, P. Greengard, D. J. Surmeier, *Science* **321**, 848–851 (2008).
12. M. Uchigashima *et al.*, *J. Neurosci.* **27**, 3663–3676 (2007).
13. V. Chevaleyre, K. A. Takahashi, P. E. Castillo, *Annu. Rev. Neurosci.* **29**, 37–76 (2006).
14. A. C. Kreitzer, R. C. Malenka, *Neuron* **60**, 543–554 (2008).
15. R. L. Albin, A. B. Young, J. B. Penney, *Trends Neurosci.* **12**, 366–375 (1989).
16. L. M. Suárez *et al.*, *Biol. Psychiatry* **75**, 711–722 (2014).
17. S. Ares-Santos, N. Granado, I. Espadas, R. Martinez-Murillo, R. Moratalla, *Neuropsychopharmacology* **39**, 1066–1080 (2014).
18. M. Gómez-Gonzalo *et al.*, *Cereb. Cortex* (2014).
19. M. Navarrete, A. Araque, *Neuron* **68**, 113–126 (2010).
20. R. Min, T. Nevian, *Nat. Neurosci.* **15**, 746–753 (2012).
21. M. Navarrete *et al.*, *PLOS Biol.* **10**, e1001259 (2012).
22. M. A. Di Castro *et al.*, *Nat. Neurosci.* **14**, 1276–1284 (2011).
23. J. Petravicz, T. A. Fiacco, K. D. McCarthy, *J. Neurosci.* **28**, 4967–4973 (2008).
24. T. Fellin *et al.*, *Neuron* **43**, 729–743 (2004).
25. M. Navarrete, A. Araque, *Neuron* **57**, 883–893 (2008).
26. G. Perea, A. Araque, *J. Neurosci.* **25**, 2192–2203 (2005).
27. G. Cui *et al.*, *Nature* **494**, 238–242 (2013).
28. A. V. Kravitz *et al.*, *Nature* **466**, 622–626 (2010).
29. I. Ruiz-DeDiego, B. Mellstrom, M. Vallejo, J. R. Naranjo, R. Moratalla, *Biol. Psychiatry* **77**, 95–105 (2015).
30. V. M. André *et al.*, *J. Neurosci.* **31**, 1170–1182 (2011).

ACKNOWLEDGMENTS

We thank D. Redish, W. Buño, A. Díez, and R. Gómez for helpful comments; J. Lerma and A. Valero for valuable help; B. Pro for technical assistance; and J. Chen (University of California–San Diego) and A. Zimmer (University of Bonn) for providing $IP_3R2^{-/-}$ and $CB1R^{-/-}$ mice, respectively. A.A. was supported by Cajal Blue Brain, the Human Frontier Science Program (grant RGP0036/2014), the Wallin family, and V. and L. Han. Support for this work was also provided by the Ministerio de Economía y Competitividad of Spain (grants BFU2013-47265-R, CSD2010-00045, and RYC-2012-12014 to G.P. and grants SAF2013-48532-R and CB06/05/0055 to R.Mo.) and by the Juan de la Cierva Program (grant JCI-2010-07693 to R.Ma.). Data described in the paper are archived at the Instituto Cajal (Madrid, Spain). The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/730/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*31*–*39*)

28 January 2015; accepted 17 June 2015
10.1126/science.aaa7945

LANGUAGE DEVELOPMENT

The developmental dynamics of marmoset monkey vocal production

D. Y. Takahashi,[1,2]* A. R. Fenley,[1,2] Y. Teramoto,[1] D. Z. Narayanan,[1,2] J. I. Borjon,[1,2] P. Holmes,[1,3] A. A. Ghazanfar[1,2,4]*

Human vocal development occurs through two parallel interactive processes that transform infant cries into more mature vocalizations, such as cooing sounds and babbling. First, natural categories of sounds change as the vocal apparatus matures. Second, parental vocal feedback sensitizes infants to certain features of those sounds, and the sounds are modified accordingly. Paradoxically, our closest living ancestors, nonhuman primates, are thought to undergo few or no production-related acoustic changes during development, and any such changes are thought to be impervious to social feedback. Using early and dense sampling, quantitative tracking of acoustic changes, and biomechanical modeling, we showed that vocalizations in infant marmoset monkeys undergo dramatic changes that cannot be solely attributed to simple consequences of growth. Using parental interaction experiments, we found that contingent parental feedback influences the rate of vocal development. These findings overturn decades-old ideas about primate vocalizations and show that marmoset monkeys are a compelling model system for early vocal development in humans.

Human vocal development is the outcome of interactions among an infant's developing body and nervous system and his or her experience with caregivers (*1*, *2*). Infant cries decline over the first 3 months as they transition into preverbal vocalizations (*3*). The rates of these transitions are influenced by social feedback: Contingent responses of caregivers spur the development of more mature vocalizations (*4*). In contrast, nonhuman primate vocalizations are widely viewed as undergoing little or no production-related acoustic changes during development, and any such changes are attributed solely to passive consequences of growth (*5*).

We tracked the vocal development of marmoset monkeys (*Callithrix jacchus*; $n = 10$)—a voluble, cooperative breeding species (*6*)—from the first postnatal day (P1) until they produced adultlike calls at 2 months of age. Recordings were taken at least twice weekly in two contexts: undirected (social isolation) and directed (with auditory, but not visual, contact with their mother or father). Such early and dense sampling is necessary to accurately capture developmental changes in marmosets because this species develops rapidly (*7*). Each recording session began with ~5 min in the undirected context followed by ~15 min in the directed context, with mothers and fathers alternating between each session. In the undirected context, infants exhibited a dramatic change in vocal production (Fig. 1A and audio S1 to S8). At P1, vocalizations were more numerous and variable in their spectrotemporal structure than those recorded in later weeks. The number and variability of calls diminished over 2 months, approaching mature vocal output with exclusive production of whistle-like "phee" calls in this context (*8*).

To quantify this developmental change as a continuous process without the bias of ethological labels (*9*), for each of the 73,421 recorded utterances, we measured four acoustic parameters similar to those used for tracking birdsong development (*10*): duration, dominant frequency, amplitude modulation (AM) frequency, and Wiener entropy (a measure of spectral flatness) (Fig. 1B). Changes in all four parameters were statistically significant ($n = 301$ sessions, $P < 0.001$), showing that vocalizations underwent a transformation in the first 2 months, whereby utterances lengthened, dominant and AM frequencies decreased, and entropy decreased. This pattern of change is consistent with both human and songbird vocal development (*10*, *11*). These changes in infant vocalizations, although not subtle, may be due solely to physical maturation (*5*). To test this, we used body weight as a proxy for overall growth [weight correlates well with vocal apparatus size in monkeys (*12*)]. Weight changes visibly contrasted with the trajectories of the acoustic parameters (Fig. 1, B and C). To quantify this difference, we used weight to predict changes in the acoustic parameters. Predicted average parameter values, given the average weight for each postnatal day, are shown in Fig. 1B. If growth completely explained the acoustic change, the residues would be uncorrelated and identically distributed across postnatal days. Using the Akaike information criterion (AIC), the best polynomial-fit order was three

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA. [2]Department of Psychology, Princeton University, Princeton, NJ 08544, USA. [3]Department of Mechanical and Aerospace Engineering and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA. [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA.
***Corresponding author. E-mail: takahashiyd@gmail.com (D.Y.T.); asifg@princeton.edu (A.A.G.)**

for all residues related to the acoustic parameters (Fig. 1D). To account for possible nonlinear relationships between growth and acoustic parameters, we log-transformed the weight and acoustic parameters. The log-transformed weight did not predict the log-transformed acoustic parameters (fig. S1, A to C). Thus, simple patterns of growth (linear or nonlinear) do not accurately predict acoustic changes in infant marmoset vocalizations.

A subset of the early vocalizations of humans and songbirds are incorporated into the adult repertoire, whereas others are transient, serving as scaffolding for later vocalizations (*3*, *10*). To test whether infant marmosets follow a similar trajectory, we first measured the extent to which their calls were distinct. Two parameters, duration and entropy, identified disjoint clusters in syllable sequences (Fig. 2A). With development, the clusters became more distinct and less numerous. Using all four parameters, we computed optimal cluster numbers for each marmoset in each session (Fig. 2B). On average, the number of clusters decreased from around four to one or two ($P < 0.001$). The clusters represent distinct ethologically based syllable types (*13*, *14*) (Fig. 2C). Phee syllables increased to over 95% of all vocalizations by 2 months ($P < 0.001$); all other calls decreased ($P = 0.005$ for trills; $P < 0.001$ for all other syllables) (Fig. 2D). The changes in syllable proportions potentially represent two independent processes: change in usage (*15*) and transformation of immature calls into mature versions (Fig. 2E). Twitters and trills are produced frequently by marmosets of all ages (*13*, *14*), but in adults, they are typically produced when in visual contact with conspecifics and not in the undirected context. Thus, twitters and trills undergo a change in usage in the first 2 months. In contrast, cries, phee-cries, and subharmonic phees are only produced by infants; mature phees are produced almost exclusively during vocal exchanges that occur when out of visual contact with conspecifics (*8*).

Because these infant-only calls share some features with the mature phee call (e.g., a common duration), we hypothesized that they represent immature phees, consistent with vocal transformations observed in preverbal human infants (*11*) and songbirds (*10*) but contrasting with previous reports on developing primates (*5*). It is possible that these transitional forms are related to growth but sound distinct because of nonlinearities in the vocal production

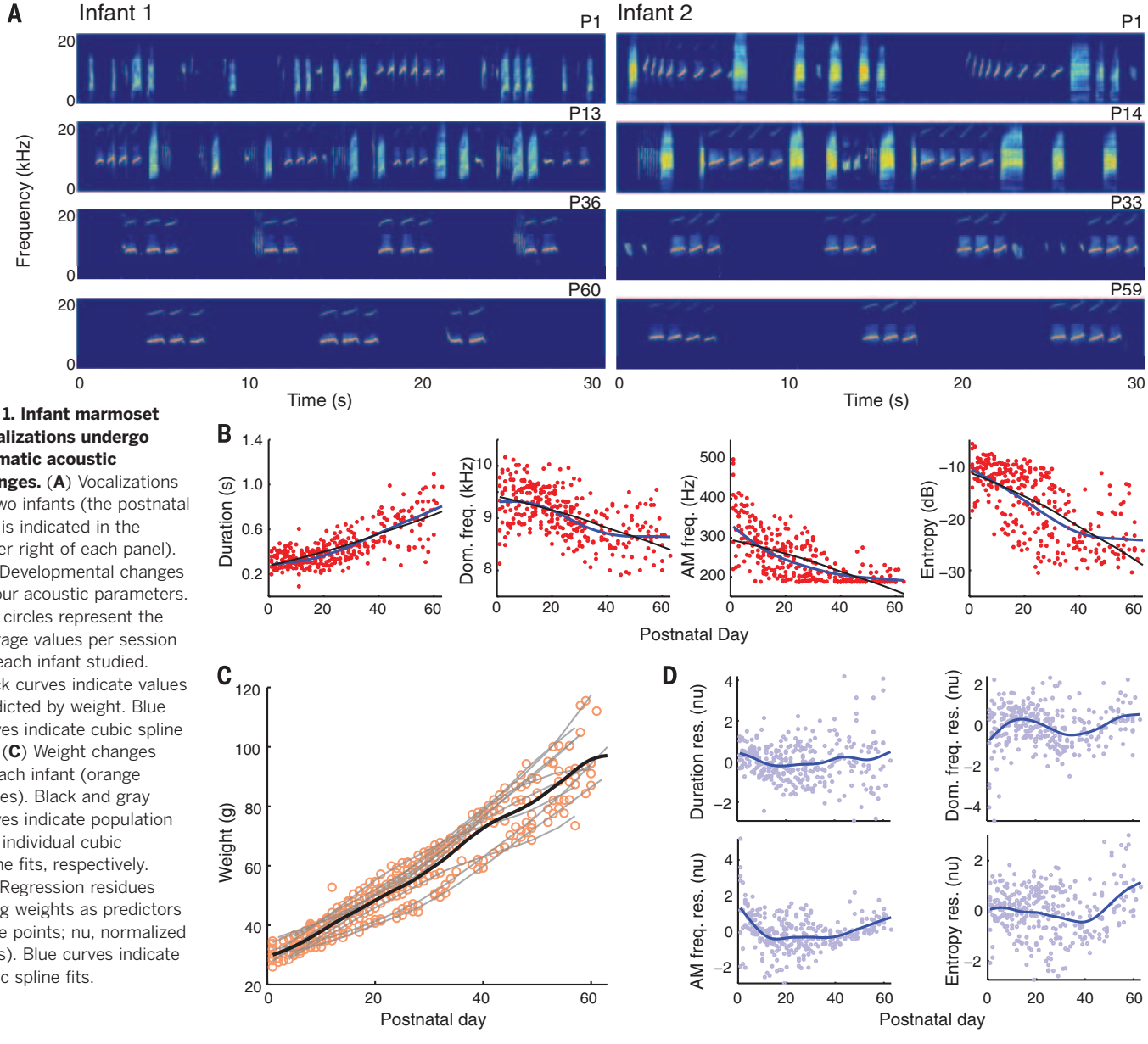


Fig. 1. Infant marmoset vocalizations undergo dramatic acoustic changes. (**A**) Vocalizations of two infants (the postnatal day is indicated in the upper right of each panel). (**B**) Developmental changes in four acoustic parameters. Red circles represent the average values per session for each infant studied. Black curves indicate values predicted by weight. Blue curves indicate cubic spline fits. (**C**) Weight changes of each infant (orange circles). Black and gray curves indicate population and individual cubic spline fits, respectively. (**D**) Regression residues using weights as predictors (blue points; nu, normalized units). Blue curves indicate cubic spline fits.

system (*16*). This would suggest that a single biomechanical mechanism generates cries, phees, and the transitional forms and that the transitional calls result from smooth changes through a parameter space. To test this idea, we developed a model based on one that successfully reproduces syllable types in zebra finch song but that is nonspecific with regard to songbird versus mammalian vocal anatomy (Fig. 3A) (*17*). Our simulations verified that the model can reproduce the marmoset call types described above (Fig. 3, B to E). The simulations also revealed the underlying biomechanics corresponding to different calls at different levels of pressure (respiratory power) and laryngeal muscle tension (Fig. 3F). Broadband cries were produced at low pressure and muscle tension, where small variations cause large changes in spectral content because of nonlinear vocal fold dynamics. Phees occurred at higher pressures and tensions, and subharmonic phees occurred in an intermediate region, supporting their classification as transitional calls. Rapid switching between high and low pressure and tension states produced the phee-cries. Throughout, linear changes in pressure and tension produced nonlinear acoustic effects.

To test the model's overall validity and the prediction that respiration during cries is less stable than during phees, we measured respiratory activity via electromyography (EMG) in five P1 infants. We investigated whether different respiratory patterns underlie cries and phees with similar intersyllable intervals (Fig. 3, G and I). The EMG signals were more uniform across phees than across cries (Fig. 3, H and J), as quantified by the cost of dynamic time warping (DTW) (*18*). For each infant, the mean DTW costs for phees were smaller than they were for cries ($P < 0.001$) (Fig. 3K). Therefore, phee syllables at least partly result from more stable respiration; immature respiratory control leads mainly to cries early in life, consistent with the model prediction. Overall, these data support our hypothesis that cries are immature phees.

Thus, although vocal acoustic changes were dramatic, physiological growth could explain the transition from cries to phees, as improved respiratory and/or laryngeal control modulates spectral parameters (Fig. 1B), reducing the entropy. However, if the cries-to-phees transition was solely driven by physical maturation, it would be impervious to social feedback. Yet, consistent with a role for vocal feedback in guiding development, marmoset monkeys exhibit a developmental pattern of *FoxP2* expression in their thalamocortical-basal ganglia circuit (*19*) that is analogous to that of songbirds and humans (*20*). This suggests that marmoset infants may use this circuit to guide their phee-call development through reward-based parental feedback, as birds and humans do (*21*). To assess the effect of parent-infant vocal interactions in marmosets, we quantified their vocal exchanges in the directed context, where infants and their mother or father were in auditory, but not visual, contact.

Infant and parent vocalizations were parsed into whole multisyllabic calls according to the

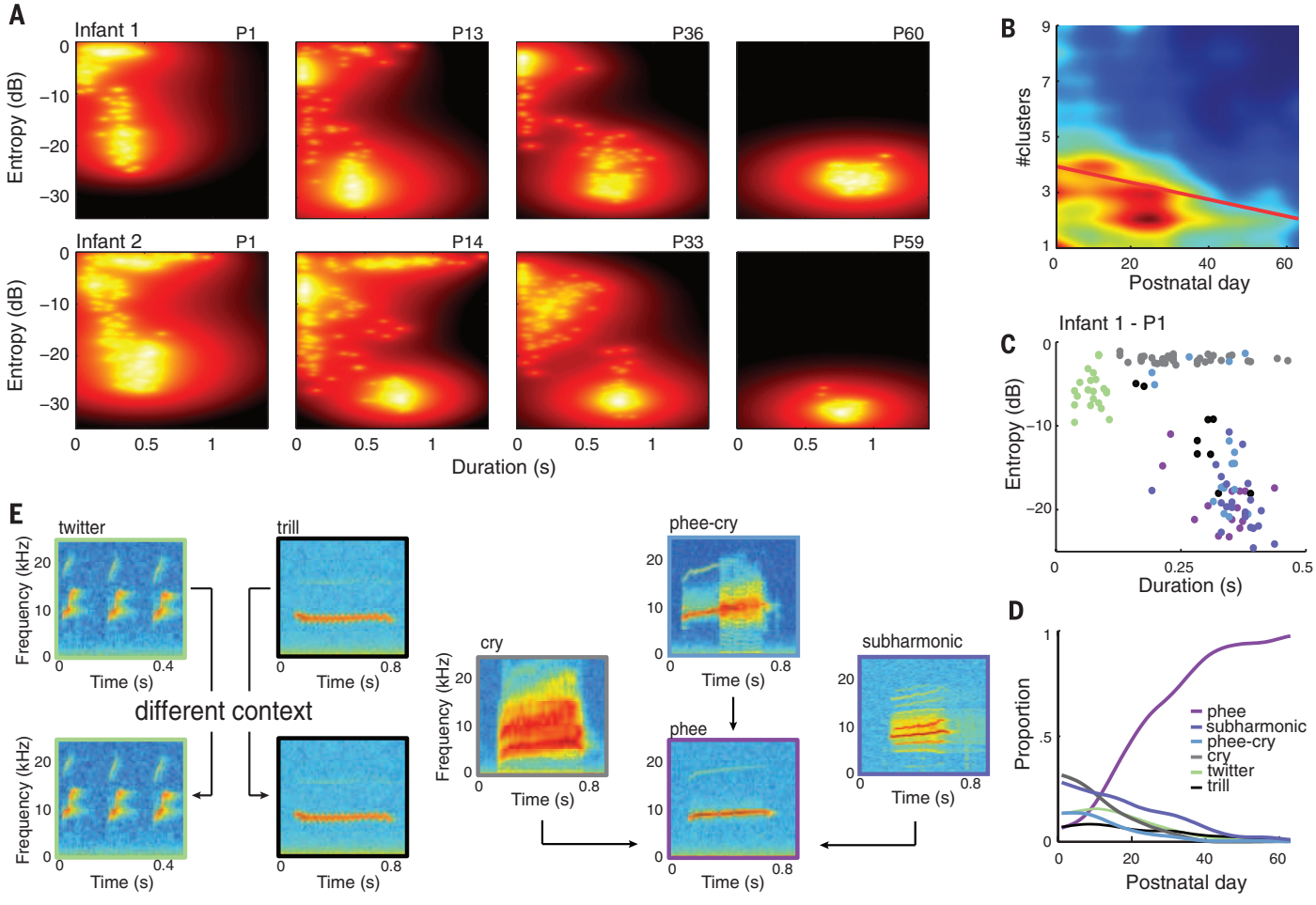


Fig. 2. A subset of infant marmoset calls transform into adultlike phee calls. (**A**) Probability maps of duration and entropy for the infants in Fig. 1A. Lighter colors indicate higher probabilities. (**B**) Distribution of cluster numbers by developmental age. Warmer colors indicate more frequent occurrences of cluster numbers. The red line is the regression fit. (**C**) Correspondence between clusters and different syllable types for a marmoset at P1. (**D**) Developmental changes in the proportion of call types. Colors in (C) indicate the same call types as in (D). (**E**) Twitters and trills change in usage, whereas cries, phee-cries, and subharmonic phees are hypothesized to transition to phee calls.

bimodal distribution of their intersyllable intervals (*8*). We recorded 8800 infant phees, 11,798 infant cries, and 6567 adult phees, of which 2512 were contingent responses to infant phees [those falling within a turn-taking interval as seen in adults (*8*)]. Parents produced mostly phee calls (>98%). Typical examples of infant phee and cry production during interactions over the first 2 months and the phee/cry ratio across days are shown in Fig. 4, A and B. As in the undirected context (Fig. 2D), cries gave way to phees, but the transition occurred rapidly. For each infant, we used the point where the phee/cry ratio first crossed zero to mark the transition day (Fig. 4C). Transitions were typically sharp, but their timing varied substantially across infants (~10 to 40 postnatal days). If physiological growth completely explained the cries-to-phees transition, the weight-change rate and the timing of the transition (zero-crossing) day would be correlated. However, we found no significant correlation (n = 10 infants, t test, P = 0.684) (Fig. 4D); growth alone cannot explain the timing of the cries-to-phees transition.

We then investigated whether parental responses to infant vocalizations affect the timing of the cries-to-phees transition. This would explain, at least partially, its variability across infants. Infants could be influenced by contingent responses only or by the total number of adult vocalizations that they hear. The fraction of infant phees that elicited contingent parental phee responses before the zero-crossing day correlated significantly with the timing of the zero-crossing day (n = 10 infants, t test, P = 0.005) (Fig. 4E). Proportions of noncontingent parental calls (91.5% of all calls on average) were not significantly correlated with this timing (n = 10 infants, t test, P = 0.558) (Fig. 4F). Therefore, contingent vocal responses from parents influence the timing of the cries-to-phees transition by reinforcing the production of phee calls.

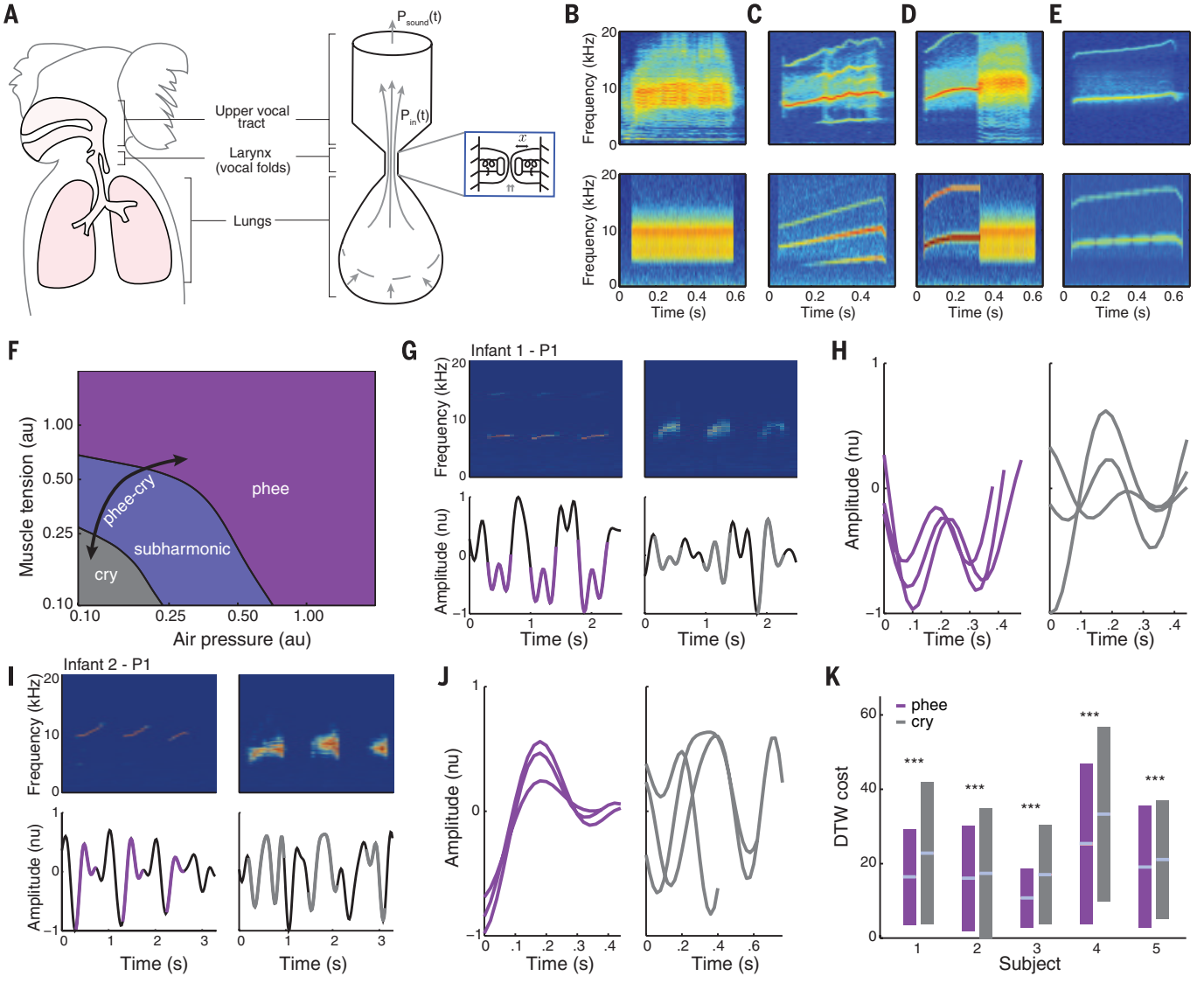


Fig. 3. Biomechanical model of vocal folds can reproduce infant marmoset calls. (**A**) Main elements of the vocal tract and respective model representations, where $P_{in}(t)$ is pressure at the inlet to the vocal tract, $P_{sound}(t)$ is pressure upon exit from the mouth, and × is the displacement of the vocal fold. Gray arrows show airflow. (**B** to **E**) Top: Representative recordings of a cry (B), subharmonic phee (C), phee-cry (D), and phee (E); bottom: corresponding model simulations. (**F**) Changing air pressure and muscle tension produces different calls (au, arbitrary units). (**G** and **I**) Top: Three-syllable phees (left) and cries (right) for two marmosets; bottom: their corresponding EMG activities. (**H** and **J**) Respiratory EMG activity during call production. Curves are the purple and gray segments from (G) and (I), aligned at the syllable onset of phees and cries, respectively. (**K**) Average DTW cost on P1 for five marmosets. Central light-colored bars show means, rectangles show standard deviations, and asterisks indicate significant differences (P < 0.001).

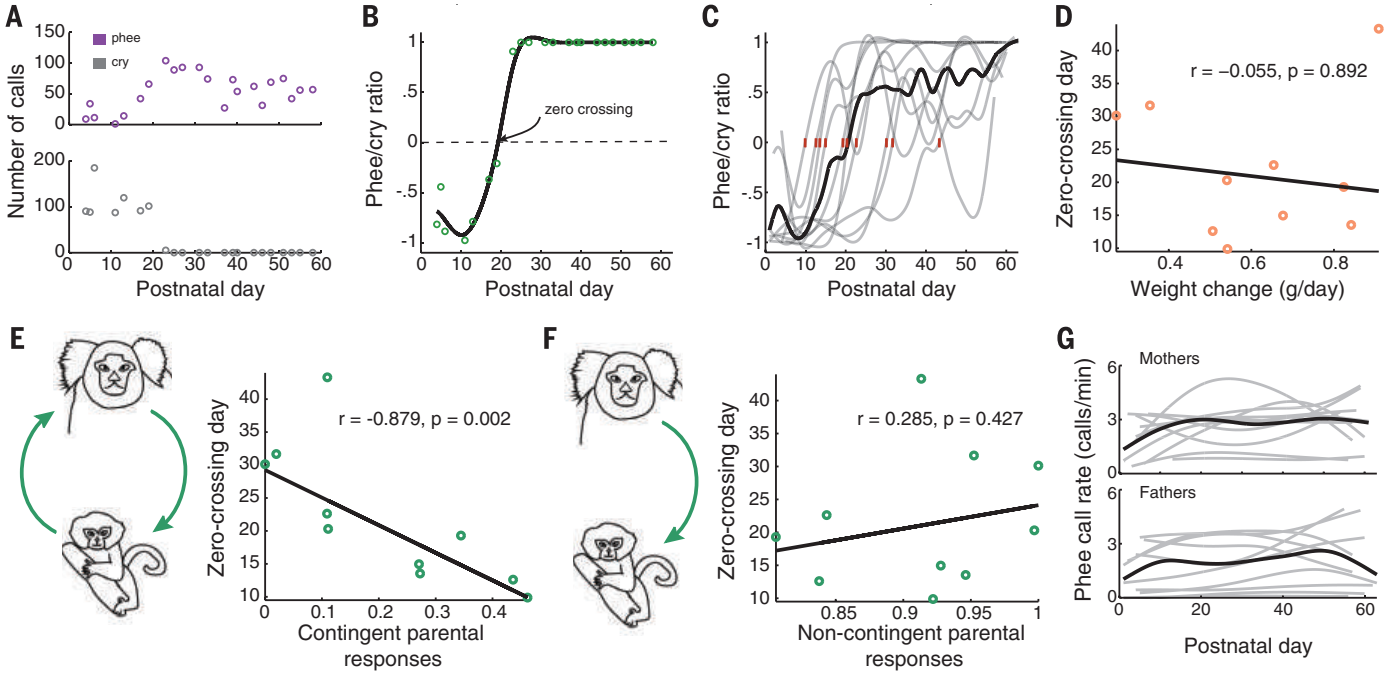


Fig. 4. Transition from cry to phee is influenced by contingent parental calls. (**A**) Numbers of cries and phees over 2 months for a single infant. (**B**) Phee/cry ratio for the infant in (A) across days. (**C**) Phee/cry ratios (gray curves) and zero-crossing days (red ticks) for each infant and for the population (black curve). Black and gray curves in (B) and (C) are cubic spline fits. (**D**) Correlation between the weight-change rate and the zero-crossing day among infants. (**E** and **F**) Correlations between the zero-crossing day and the proportion of contingent and noncontingent parental responses, respectively. (**G**) Rates of individual parental phee-call production during infant development (gray) and the population average (black).

We address two possible caveats to this conclusion. First, it is possible that, through shared genetics, fast-transitioning infants are born to more vocally interactive parents. To test this, we correlated the frequency of contingent parental calls and the zero-crossing day for six full siblings born from the same parents. If shared genetics were driving the result, then there would be no correlation between contingent parental responses and the zero-crossing day. We found, however, that there remained a statistically significant correlation (n = 6 infants, P = 0.046) (fig. S2). Moreover, we found no difference between the slopes of the regressions for the full-siblings and all-infants data (test for equality, P = 0.953).

Second, it is possible that changing patterns of infant calling are due to changes in parental call output. The phee-call production rates of each infant's parents during development are shown in Fig. 4G; neither parent changed their production rates (mother, P = 0.132; father, P = 0.235). Based on these analyses, we conclude that the cries-to-phees transition is influenced by contingent responses from parents, not by shared genetics or changes in parental vocal output.

Our findings demonstrate that infant marmoset calls undergo dramatic changes during the first 2 months of life, transforming from cries into mature, adultlike phee calls. The timing of this transition is partly attributable to maturation but is also influenced by contingent parental vocal feedback. This is consistent with preverbal vocal development in humans, whereby (i) natural categories of sounds change as respiratory, laryngeal, and facial components mature, and (ii) in parallel, vocal feedback sensitizes infants to certain features of those sounds, and the sounds are modified accordingly. Our findings contrast with previous reports that nonhuman primate vocalizations undergo little or no postnatal change and are impervious to social feedback (*5*). The complex and socially dependent vocal development we observed in marmoset monkeys may be a necessary condition of the vocal learning observed in humans.

REFERENCES AND NOTES

1. E. Thelen, in *Biological and Behavioral Determinants of Language Development*, N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, M. Studdert-Kennedy, Eds. (Lawrence Erlbaum Associates, Hillsdale, NJ, 1991), pp. 339–362.
2. L. Byrge, O. Sporns, L. B. Smith, *Trends Cogn. Sci.* **18**, 395–403 (2014).
3. D. K. Oller, *The Emergence of the Speech Capacity* (Lawrence Erlbaum Associates, Mahwah, NJ, 2000).
4. M. H. Goldstein, A. P. King, M. J. West, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8030–8035 (2003).
5. S. E. R. Egnor, M. D. Hauser, *Trends Neurosci.* **27**, 649–654 (2004).
6. J. I. Borjon, A. A. Ghazanfar, *Brain Behav. Evol.* **84**, 93–102 (2014).
7. A. de Castro Leão, A. Duarte Dória Neto, M. B. C. de Sousa, *Comput. Biol. Med.* **39**, 853–859 (2009).
8. D. Y. Takahashi, D. Z. Narayanan, A. A. Ghazanfar, *Curr. Biol.* **23**, 2162–2168 (2013).
9. D. Lipkind, O. Tchernichovski, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15572–15579 (2011).
10. O. Tchernichovski, P. P. Mitra, T. Lints, F. Nottebohm, *Science* **291**, 2564–2569 (2001).
11. E. Scheiner, K. Hammerschmidt, U. Jurgens, P. Zwirner, *J. Voice* **16**, 509–529 (2002).
12. W. T. Fitch, *J. Acoust. Soc. Am.* **102**, 1213–1222 (1997).
13. B. M. Bezerra, A. Souto, *Int. J. Primatol.* **29**, 671–701 (2008).
14. A. L. Pistorio, B. Vintch, X. Wang, *J. Acoust. Soc. Am.* **120**, 1655–1670 (2006).
15. A. M. Elowson, C. T. Snowdon, C. Lazaro-Perea, *Behaviour* **135**, 643–664 (1998).
16. W. T. Fitch, J. Neubauer, H. Herzel, *Anim. Behav.* **63**, 407–418 (2002).
17. A. Amador, Y. S. Perl, G. B. Mindlin, D. Margoliash, *Nature* **495**, 59–64 (2013).
18. J. A. Kogan, D. Margoliash, *J. Acoust. Soc. Am.* **103**, 2185–2196 (1998).
19. M. Kato *et al.*, *Brain Lang.* **133**, 26–38 (2014).
20. I. Teramitsu, L. C. Kudo, S. E. London, D. H. Geschwind, S. A. White, *J. Neurosci.* **24**, 3152–3163 (2004).
21. S. Syal, B. L. Finlay, *Dev. Sci.* **14**, 417–430 (2011).

ACKNOWLEDGMENTS

We thank L. Kelly for providing comments on an earlier draft. This work was supported by a Pew Latin American Fellowship (D.Y.T.), a Brazilian Science Without Borders Fellowship (D.Y.T.), an NSF Graduate Fellowship (J.I.B.), and the James S. McDonnell Scholar Award (A.A.G). Data for each figure are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/734/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 and S2
References (*22*–*29*)
Audio S1 to S8
Supplementary Data

13 March 2015; accepted 12 June 2015
10.1126/science.aab1058

EBOLA VACCINE

VSV-EBOV rapidly protects macaques against infection with the 2014/15 Ebola virus outbreak strain

Andrea Marzi,[1] Shelly J. Robertson,[1] Elaine Haddock,[1] Friederike Feldmann,[2] Patrick W. Hanley,[2] Dana P. Scott,[2] James E. Strong,[3] Gary Kobinger,[3] Sonja M. Best,[1] Heinz Feldmann[1]*

The latest Ebola virus (EBOV) epidemic spread rapidly through Guinea, Sierra Leone, and Liberia, creating a global public health crisis and accelerating the assessment of experimental therapeutics and vaccines in clinical trials. One of those vaccines is based on recombinant vesicular stomatitis virus expressing the EBOV glycoprotein (VSV-EBOV), a live-attenuated vector with marked preclinical efficacy. Here, we provide the preclinical proof that VSV-EBOV completely protects macaques against lethal challenge with the West African EBOV-Makona strain. Complete and partial protection was achieved with a single dose given as late as 7 and 3 days before challenge, respectively. This indicates that VSV-EBOV may protect humans against EBOV infections in West Africa with relatively short time to immunity, promoting its use for immediate public health responses.

The largest documented outbreak of Ebola hemorrhagic fever (EHF) caused by the Ebola virus (EBOV) started in rural Guinea in late 2013 and spread regionally, predominantly affecting Guinea, Sierra Leone, and Liberia (*1*, *2*). This global health crisis to date accounts for nearly 27,200 cases and more than 11,100 deaths, with a high infection rate among health care workers (*2*). The global scale of this outbreak is further demonstrated by the impact on public health systems in multiple countries that had to respond to introduction of cases, including evacuations of exposed and/or infected aid workers (*2*). In addition, the inability to recruit timely and adequate medical, logistic, and financial support became a serious problem. Thus, this outbreak emphasizes the urgent need to develop therapeutics and vaccines for filoviruses to avert future outbreaks, and several promising experimental approaches have been accelerated for clinical trials (*3*). One of the most promising experimental vaccines is based on the recombinant vesicular stomatitis virus platform expressing the EBOV glycoprotein (GP) (previously referred to as rVSV-ZEBOV, here designated VSV-EBOV) as an immunogen. VSV-EBOV is a live-attenuated vaccine vector that has demonstrated an adequate preclinical safety profile and shown marked efficacy in pre- and postexposure vaccination in rodent and macaque models (*4*, *5*). The mechanism of protection for prophylactic vaccination is thought to be largely antibody-mediated (*6*).

VSV-EBOV was recently evaluated in phase 1 clinical trials in humans at several worldwide locations. Despite reports of arthritis in about 20% of volunteers at a single study site that used the highest vaccine dose, the vaccine was found in general to be safe and immunogenic, warranting further evaluation for efficacy (*7*, *8*). Unexpectedly, preclinical efficacy data for this or any other EBOV vaccine currently in phase 1 clinical trials against the current West African EBOV-Makona strain is lacking. Because genetic divergence among known EBOV strains is similar, with about 3% whole-genome divergence (*1*), one may assume that VSV-EBOV will protect against all

[1]Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA. [2]Rocky Mountain Veterinary Branch, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA. [3]Special Pathogens Program, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada.
*Corresponding author. E-mail: feldmannh@niaid.nih.gov

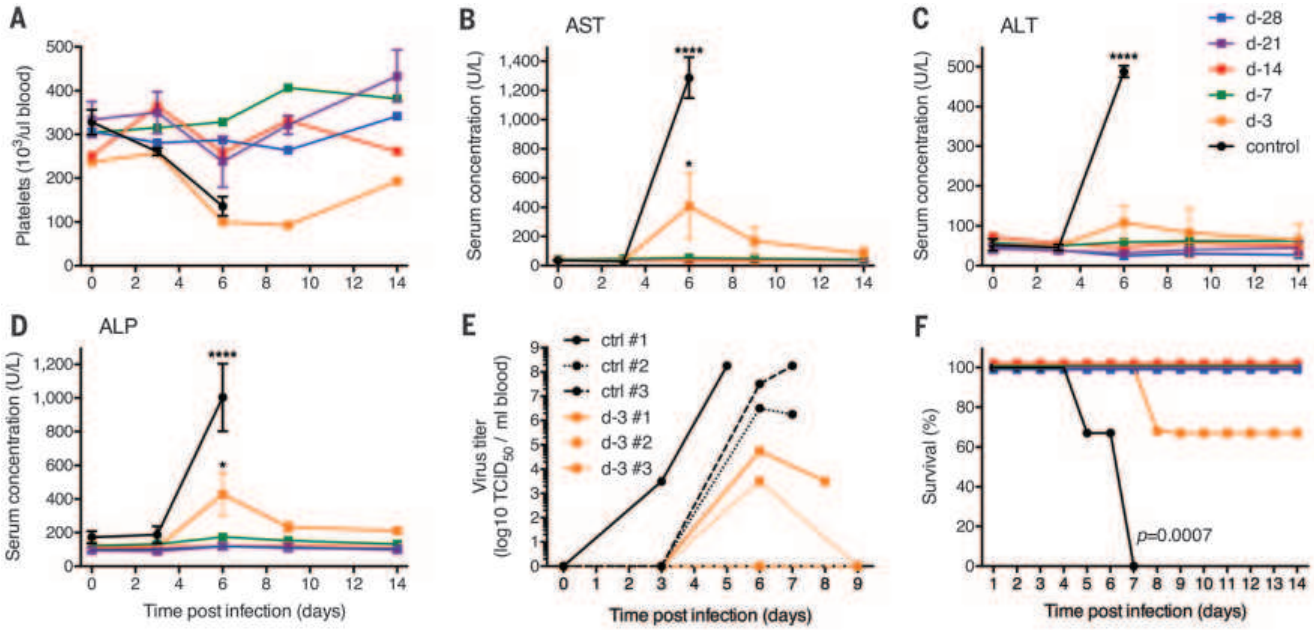


Fig. 1. Clinical parameters of infected NHPs. (**A**) Platelet counts of infected NHPs in EDTA blood at every examination day. Concentrations of aspartate aminotransferase (AST) (**B**), alanine transaminase (ALT) (**C**), and alkaline phosphatase (ALP) (**D**) were determined in serum samples collected on examination days (d) after EBOV-Makona challenge. U, units. Two-way analysis of variance (ANOVA) with Tukey multiple comparison post test was used to determine statistical significance at the level of 0.05 (*) and 0.0001 (****) for data presented in (B) to (D). Error bars indicate SD. (**E**) Viremia (titers) for individual animals on examination days and time of euthanasia. $TCID_{50}$, median tissure culture infectious dose. (**F**) Survival curves showing outcome of the different vaccine groups after challenge. Statistical significance was assessed by using the Kaplan-Meier method.

known EBOV strains. However, analysis of sequence data from May and June 2014 in Sierra Leone suggested a higher mutation rate of EBOV in this ongoing outbreak (*9*), raising concerns about the applicability of current experimental intervention strategies, including VSV-EBOV, against the emerged EBOV-Makona strain. More-recent sequence data from Mali (October and November 2014) indicated a mutation rate of EBOV in West Africa similar to observations in past outbreaks (*10*), lessening those concerns. However, to address this important question, we performed efficacy testing of good manufacturing practice (GMP)–grade VSV-EBOV in a recently established cynomolgus macaque challenge model for the West African EBOV-Makona strain (*11*). In parallel, we investigated the time to protective immunity after a single high-dose VSV-EBOV vaccination against challenge with EBOV-Makona. All infectious animal work was performed in the maximum containment laboratory at the Rocky Mountain Laboratories, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA, applying standard operating protocols approved by the Institutional Biosafety Committee (*12*).

Fifteen cynomolgus macaques were randomly assigned to groups of two or three animals and immunized with a single intramuscular injection of 5×10^7 plaque-forming units (PFU) of GMP-grade VSV-EBOV at 28, 21, 14, 7, or 3 days before challenge. This vaccine dose was chosen to be equivalent to the highest immunization dose evaluated in the recent phase 1 clinical trials (*7*). The control group (three animals) was immunized 28 days before challenge by the same route and dose with the VSV-Marburg virus vaccine (VSV-MARV), previously shown not to protect against EBOV challenge (*5*, *6*). Daily monitoring, including weekly blood sampling and clinical examinations of the animals, revealed normal behavior and no detectable adverse effects after vaccination. Challenge was performed with an intramuscular injection into the opposite leg of 1000 PFU of EBOV-Makona, established as a lethal dose (*11*). Health monitoring and clinical examinations were intensified and continued until the study ended 42 days after challenge or until euthanasia was required owing to the severity of disease. Humane end-point criteria, specified and approved by the Institutional Animal Care and Use Committee, were applied to determine when animals should be humanely euthanized.

The VSV-MARV–vaccinated control animals and one animal in the day-3 vaccination group [nonhuman primate (NHP) d-3 number 1] developed classical severe EHF with macular cutaneous rash, thrombocytopenia (Fig. 1A), increased liver enzyme levels (Fig. 1, B to D), and viremia (Fig. 1E); these animals had to be euthanized on days 5, 7, 7, and 8 after challenge (Fig. 1F). Another animal in the day-3 vaccination group (NHP d-3 number 2) developed moderate signs of EHF, including viremia (Fig. 1E) and mild rash, but did not reach the clinical score for humane end-point euthanasia, cleared the virus by day 9 (Fig. 1E), and survived. The third animal in this group developed only very mild signs of disease without viremia or rash and recovered quickly. All nine remaining animals in the day-28, -21, -14, and -7 vaccination groups did not develop any clinical signs of disease, showed no EBOV viremia, and did not have abnormal changes in hematology or liver enzymes, indicating complete protection against lethal EBOV-Makona challenge (Fig. 1).

Previous work had established antibody responses as the main mechanism of protection for prophylactic use of VSV-EBOV in macaques (*6*). The role of cellular immune responses for protection with this vaccine has never been clearly established, and $CD4^+$ and $CD8^+$ T cell responses are nearly undetectable during immunization and at time of challenge (*4*, *6*). Therefore, we focused on antibody responses in the vaccinated animals. At the time of challenge, animals in the day-28, -21, and -14 vaccination groups showed potent EBOV-GP–specific responses with immunoglobulin (IgG) titers $>10^4$ (range of 51,200 to 102,400) (Fig. 2A and fig. S1A). EBOV-GP–specific antibody titers $>10^4$ were also associated with protection during testing of the adenovirus 5 vaccine (rAd5-EBOV) platform (*13*, *14*), as well as for a second-generation VSV-based vector not yet in phase 1 clinical trials (*15*). In contrast, titers of animals in the day-7 and -3 vaccination groups approximated background level at the time of challenge (Fig. 2A and fig. S1A). As expected, the VSV-MARV–vaccinated control animals did not develop any EBOV-GP–specific IgG responses (Fig. 2A). In vitro titers of neutralizing antibody after intramuscular VSV-EBOV immunization have previously been shown to reach relatively low levels before challenge (*5*, *6*, *16*). In this study, only the day-28, day-21, and day-14 vaccinated animals had neutralizing antibodies at the time of challenge, with similar titers against three distinct EBOV strains: EBOV-Makona, the challenge virus (Fig. 2B, bar graph); EBOV-Kikwit, the source of the vaccine immunogen (Fig. 2C, bar graph); and EBOV-Mayinga, the prototype of the species

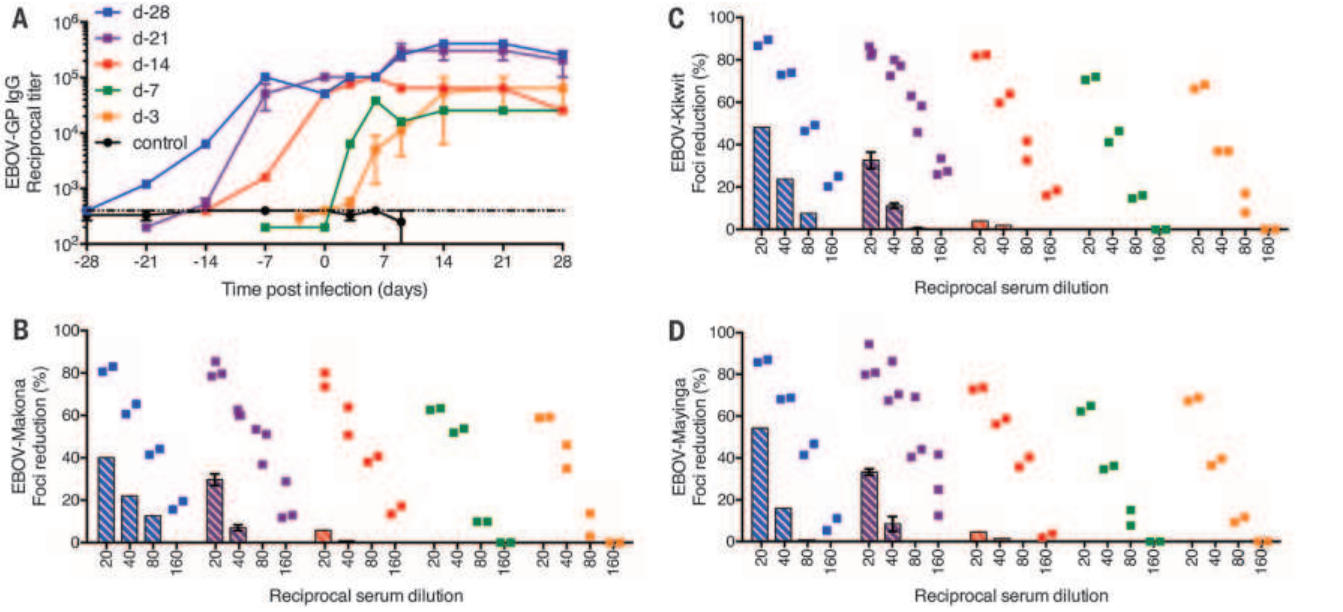


Fig. 2. Humoral immune responses of NHPs. EBOV-GP–specific IgG titers were determined in the serum of animals at examination days. (**A**) Summary of EBOV-GP-IgG titers of all groups. The dotted line marks the enzyme-linked immunosorbent assay (ELISA) cut-off based on titers obtained from negative control animals. Neutralizing antibody titers were determined by using the focus reduction assay against three distinct EBOV strains: (**B**) EBOV-Makona, the challenge strain; (**C**) EBOV-Kikwit, the source of the vaccine immunogen; and (**D**) EBOV-Mayinga, the prototype strain of the species. Data are presented as percent foci reduction in relation to a negative control serum. Bar graphs (average of animals per group) represent the neutralizing activity on day 0 (day of challenge); individual animal serum-neutralizing activity is shown for day 42 after challenge.

(Fig. 2D, bar graph). For the day-28 group, the neutralizing activity of sera was similar to that reported previously (*5*, *6*, *16*). However, increased neutralizing antibody titers were measured in serum of all surviving animals 42 days after EBOV-Makona infection (Fig. 2, B to D), a phenomenon that has been reported previously in response to this vaccine (*6*). There was no significant difference in the neutralizing activity of the day-0 and -42 serum samples against the three distinct EBOV strains (EBOV-Makona, EBOV-Kikwit, and EBOV-Mayinga) (Fig. 2, B to D). These results indicate that the VSV-EBOV vaccine will be protective with comparable efficacy against different EBOV strains, including the currently circulating West African strain.

We also measured systemic levels of various cytokines, chemokines, and other soluble factors in the serum of infected animals. An aberrant cytokine and chemokine response is considered a hallmark of severe or lethal EHF in the macaque model and in humans (*11*, *17*). Consistent with these reports, all VSV-MARV–vaccinated control animals showed marked increases of interleukin-1β (IL-1β), IL-6, IL-15, interferon-γ (IFN-γ), IL-10, and monocyte chemotactic protein (MCP)-1 after EBOV-Makona challenge, indicative of a cytokine storm (Fig. 3). In contrast, VSV-EBOV–vaccinated animals (day-28, -21, -14, and -7 groups) showed low to undetectable levels of these cytokines and chemokines, reflecting effective control of EBOV replication (Fig. 3). The animals in the day-3 vaccination group displayed intermediate levels of these cytokines and chemokines on days 3, 6, and 9 after challenge (Fig. 3). The one animal that succumbed to challenge in this group showed similar cytokine or chemokine levels as the control animals, whereas both animals that survived in the day-3 group converted on day 14 after challenge to levels of cytokines and chemokines seen in all other vaccinated and protected animals (Fig. 3).

Protection of animals vaccinated 7 and 3 days before challenge occurred despite the lack of detectable EBOV-GP–specific IgG at time of challenge (Fig. 2A). We therefore tested the sera of the VSV-MARV–vaccinated control and the day-7 and day-3 vaccinated animals for EBOV-GP–specific IgM responses. Vaccine-specific IgM antibodies at the time of challenge were undetectable for the day-3 vaccinated animals and weakly positive for the day-7 vaccinated animals (fig. S1B). EBOV-GP–specific IgM responses developed in both vaccine groups within a week after vaccination and increased in titer over time (fig. S1B). This observation, combined with the elevated cytokine and chemokine profiles, prompted us to analyze further early innate responses to better

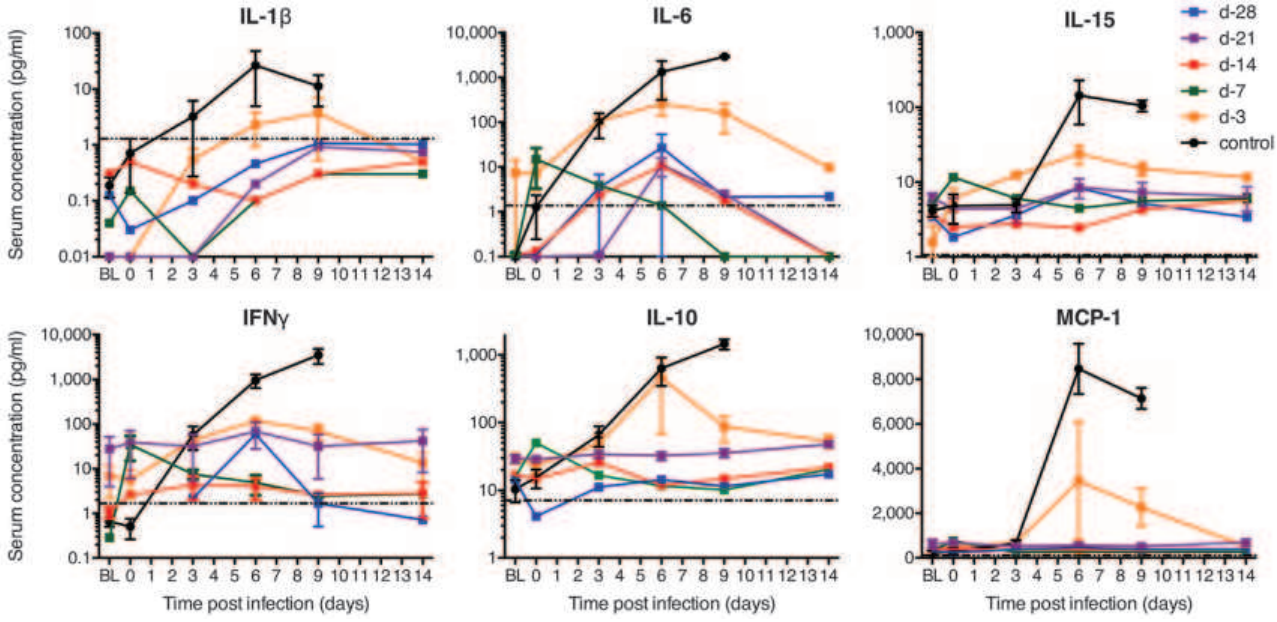


Fig. 3. Serum cytokine and chemokine levels of NHPs. Kinetics of IL-1β, IL-6, IL-15, IFN-γ, IL-10, and MCP-1 were analyzed in serum samples of each animal collected on examination days and time of euthanasia. The dotted line in each panel marks the assay detection limit. Terminal values for the control animal euthanized on day 5 were added to the day-6 time point in this group; similarly, the day-9 time point in the control and day-3 vaccination groups represents all values, including terminal samples obtained from NHPs euthanized on days 7 and 8. BL indicates baseline (before vaccination).

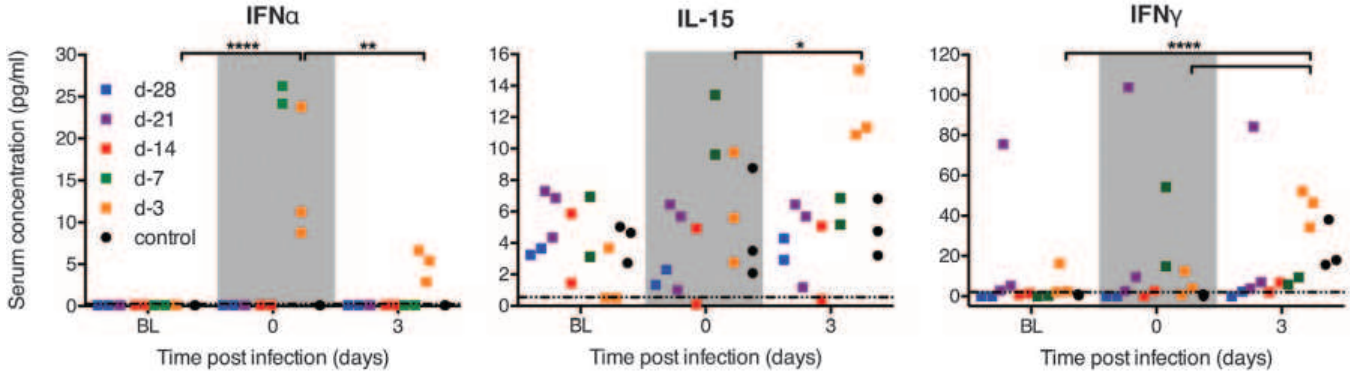


Fig. 4. Serum cytokine levels early after challenge. Serum concentrations for IFN-α, IL-15, and IFN-γ were determined before vaccination (baseline, BL), on day of challenge (gray region, day 0), and early after EBOV-Makona challenge (day 3). Each square represents an individual animal at the indicated time point. The dotted line in each panel marks the assay detection limit. Two-way ANOVA with Tukey multiple comparison post test was used to determine statistical significance at the level of 0.05 (*), 0.01 (**), and 0.0001 (****) for the day-3 vaccination group.

understand the mechanism by which these animals, particularly the day-3 vaccinated NHPs, were protected. VSV is a potent inducer of type I IFN responses (*18*, *19*), and transient, low-level VSV viremia is known to be associated with VSV-EBOV vaccination in NHPs and humans (*5*, *7*, *8*, *20*, *21*). With the exception of the day-3 vaccinated animals that had VSV-specific RNA levels of 1 to 6 PFU equivalents, all vaccinated NHPs had cleared VSV at the time of challenge (day 0). The day-3 vaccinated animals expressed higher levels of IFN-α at both day 0 and day 3 after challenge compared with the other groups (Fig. 4). Elevated IFN-α was associated with increased IL-15 and IFN-γ at day 3 after challenge (Fig. 4). Day-7 vaccinated animals also showed elevated levels of IFN-α, IL-15, and IFN-γ at time of challenge that diminished by day 3 (Fig. 4). This cytokine signature is highly suggestive of natural killer (NK) cell activation, which requires type I IFN and IL-15 production by activated macrophages, resulting in high levels of IFN-γ expression by NK cells and direct killing of virus-infected cells. More comprehensive comparison of serum cytokines in all vaccinated groups revealed up-regulation of other potentially macrophage-derived cytokines (IL-6, transforming growth factor–α, and MCP-1) in day-3 vaccinated animals by 3 days after challenge (fig. S2). These findings suggest that VSV vaccination induced activation of macrophages and NK cells, the latter of which has previously been implicated in survival from EBOV infections (*22*, *23*). Although EBOV-VP35 and -VP24 proteins have potent IFN antagonist activity (*24*, *25*), induction of innate immunity by VSV vaccination likely establishes an early antiviral state in the host that limits EBOV replication upon challenge (Fig. 1E, day 6). However, it is unlikely that innate immune responses alone can protect from lethal EBOV infection, because VSV-MARV–vaccinated control animals also showed elevated levels of IFN-γ on day 3 after challenge, likely triggered by EBOV-Makona infection, and were not protected (Fig. 4). In contrast to these control animals, the day-3 immunized animals had measurable EBOV-GP–specific IgM and IgG antibodies by days 3 and 6 after challenge, respectively, leading to protection (Fig. 2A and fig. S1). Thus, innate immune activation by VSV-EBOV may provide a window of protection that limits virus replication in the critical period needed for the development of specific adaptive responses, most importantly antibodies. Future studies will specifically compare naïve versus control vaccinated versus VSV-EBOV–vaccinated animals, including detailed analysis of innate and adaptive immune responses.

This study provides the preclinical efficacy data for GMP-grade VSV-EBOV in the gold standard macaque disease model for EHF and thus fills a gap in our understanding of appropriate responses to West African EHF. Vaccination with a single high dose of GMP-grade VSV-EBOV (previously rVSV-ZEBOV), the same preparation that was recently evaluated in phase 1 clinical trials and is currently being administered in vaccine trials in West Africa, did not result in any adverse effects in macaques, which largely is in line with reports of minor adverse effects in human volunteers after VSV-EBOV vaccination at phase 1 clinical study sites in Europe, Africa, and the Americas (*7*, *8*). Total EBOV-GP–specific IgG titers >10^4 seem to be a correlate of protection for prophylactic use of VSV-EBOV (*6*), similar to the rAd5-EBOV (*13*, *14*) and a second-generation rVSV/ZEBOV (*15*) vaccine platform. However, the time to immunity for VSV-EBOV is short, providing partial protection against lethal disease even when administered 3 days before challenge. The initial mechanism here appears to be a strong innate immune response to VSV-EBOV vaccination that may include macrophage activation and NK cell–mediated control of EBOV replication before specific adaptive responses develop. Thus, our data warrant the continuation of safety and immunogenicity trials with GMP-grade VSV-EBOV in humans and the continuation of the recently initiated clinical trials in West Africa (*26*), in particular the ring vaccination trial in Guinea (*27*).

REFERENCES AND NOTES

1. S. Baize *et al.*, *N. Engl. J. Med.* **371**, 1418–1425 (2014).
2. World Health Organization (WHO), "Ebola situation report, 3 June 2015," http://apps.who.int/ebola/ebola-situation-reports (2015).
3. U.S. National Institutes of Health, www.ClinicalTrials.gov (2015).
4. T. W. Geisbert, H. Feldmann, *J. Infect. Dis.* **204** (suppl. 3), S1075–S1081 (2011).
5. S. M. Jones *et al.*, *Nat. Med.* **11**, 786–790 (2005).
6. A. Marzi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1893–1898 (2013).
7. S. T. Agnandji *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1502924 (2015).
8. J. A. Regules *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa1414216 (2015).
9. S. K. Gire *et al.*, *Science* **345**, 1369–1372 (2014).
10. T. Hoenen *et al.*, *Science* **348**, 117–119 (2015).
11. A. Marzi *et al.*, *Emerg. Infect. Dis.* 10.3201/eid2110.150259 (2015).
12. See supplementary materials on *Science* Online.
13. N. J. Sullivan *et al.*, *Nature* **408**, 605–609 (2000).
14. N. J. Sullivan *et al.*, *PLOS Med.* **3**, e177 (2006).
15. C. E. Mire *et al.*, *Nature* **520**, 688–691 (2015).
16. X. Qiu *et al.*, *PLOS ONE* **4**, e5547 (2009).
17. H. Feldmann, T. W. Geisbert, *Lancet* **377**, 849–862 (2011).
18. A. Roberts *et al.*, *J. Virol.* **72**, 4704–4711 (1998).
19. A. Roberts *et al.*, *J. Virol.* **73**, 3723–3732 (1999).
20. L. Lai *et al.*, *JAMA* **313**, 1249–1255 (2015).
21. S. Günther *et al.*, *J. Infect. Dis.* **204** (suppl. 3), S785–S790 (2011).
22. H. Feldmann *et al.*, *PLOS Pathog.* **3**, e2 (2007).
23. K. L. Warfield *et al.*, *J. Exp. Med.* **200**, 169–179 (2004).
24. C. F. Basler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12289–12294 (2000).
25. M. Mateo *et al.*, *J. Virol.* **84**, 1169–1175 (2010).
26. U.S. National Institutes of Health, www.clinicaltrials.gov/ct2/show/NCT02344407?term=NCT02344407&rank=1 and www.clinicaltrials.gov/ct2/show/NCT02378753?term=NCT02378753&rank=1 (2015).
27. WHO, "Ebola vaccine efficacy trial to launch in Guinea," www.who.int/mediacentre/news/releases/2015/ebola-vaccine-trial/en/ (2015).

ACKNOWLEDGMENTS

We thank the Rocky Mountain Veterinary Branch [NIH, National Institute of Allergy and Infectious Diseases (NIAID)] for its support of this study. This work was jointly funded by the Division of Intramural Research, NIAID, NIH; the Public Health Agency of Canada; and the Center for Security Science Program, Canada. Use of the GMP-produced vaccine material is regulated through the Public Health Agency of Canada. The authors declare no competing financial interests. H.F. is a co-applicant on two patents regarding VSV-EBOV (U.S. patent application nos. PCT/CA03/01125US and 61/014,626). The data are available via figshare: http://dx.doi.org/10.6084/m9.figshare.1456239.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/739/suppl/DC1
Materials and Methods
Figs. S1 and S2
Reference (*28*)

20 April 2015; accepted 29 June 2015
Published online 6 August 2015
10.1126/science.aab3920

DNA REPAIR

Mus81 and converging forks limit the mutagenicity of replication fork breakage

Ryan Mayle,[1] Ian M. Campbell,[1] Christine R. Beck,[1] Yang Yu,[1] Marenda Wilson,[1]* Chad A. Shaw,[1] Lotte Bjergbaek,[2] James R. Lupski,[1,3,4] Grzegorz Ira[1]†

Most spontaneous DNA double-strand breaks (DSBs) result from replication-fork breakage. Break-induced replication (BIR), a genome rearrangement–prone repair mechanism that requires the Pol32/POLD3 subunit of eukaryotic DNA Polδ, was proposed to repair broken forks, but how genome destabilization is avoided was unknown. We show that broken fork repair initially uses error-prone Pol32-dependent synthesis, but that mutagenic synthesis is limited to within a few kilobases from the break by Mus81 endonuclease and a converging fork. Mus81 suppresses template switches between both homologous sequences and diverged human *Alu* repetitive elements, highlighting its importance for stability of highly repetitive genomes. We propose that lack of a timely converging fork or Mus81 may propel genome instability observed in cancer.

Template switches and mutation clusters of <1 to tens of kilobases occur at double-strand breaks (DSBs) repaired by homologous recombination (HR) and are implicated in rapid evolution, adaptation, and tumorigenesis (*1*–*4*). We uncovered a mechanism that restricts error-prone synthesis during HR at the most common type of DSB, broken replication forks. The break-induced replication (BIR) pathway has been proposed for repair of single-ended

DSBs (seDSBs), but its contribution to broken fork repair is unknown. In BIR, a displacement loop (D-loop), the initial recombination intermediate, is extended by Polδ with Pol32 and Pif1 helicase (*5*–*7*). BIR is highly mutagenic and prone to template switches (*2*, *3*), likely because of long single-strand DNA intermediates and instability of the D-loop (*6*, *8*). A similar pathway exists in humans (*9*, *10*). Here, we show that Mus81 and converging replication forks arriving from the opposite direction alleviate much of the genome-destabilizing consequences of BIR.

To understand the contribution of mutagenic Pol32–mediated BIR to broken fork repair, we used a galactose-inducible, nick-mediated, recombination assay (*11*). In this system, Flp1H305L, a step arrest mutant expressed from an inducible *GAL10* promoter, creates a long-lived single-stranded break at an FRT (flippase recognition target) site, which is converted to a seDSB when encountered by a replication fork. To ascertain the potential contribution of converging forks limiting the extent of repair-specific DNA synthesis, the FRT site was inserted either between two efficient origins of replication on ChrVI or at subtelomeric regions on ChrII or ChrIV, where only inefficient or dormant origins are present distal to the FRT (*12*, *13*). We confirmed that cells predominantly induce seDSBs originating from one side of the nick (Fig. 1A and fig. S1).

We first determined whether proteins important for BIR—Pif1 and Pol32—are required for efficient repair of broken replication forks. Wild-type and mutant cells were plated on media containing either glucose or galactose. Both *pif1*-m2 and *pol32*Δ cells grow equivalently to wild type, even when fork breakage occurs between an efficient origin of replication and the telomere (Fig. 1B). In agreement with the nonessential role of Pol32 and Pif1 in repair of broken forks, *pif1*-m2 and *pol32*Δ cells are not sensitive to camptothecin (CPT), a topoisomerase I inhibitor, which induces DNA nicks that can be converted to broken forks during replication (fig. S2).

In BIR, studied outside the context of replication, D-loop migration results in conservative inheritance of newly synthesized strands (*6*, *14*). Consequently, structure-specific endonucleases Mus81, which can convert a D-loop to a replication fork (*15*, *16*), and Yen1 are mostly dispensable (*7*, *17*). In contrast, elimination of Mus81 and Yen1 leads to a dramatic defect in broken fork repair (Fig. 1B), which is consistent with previous observations in budding and fission yeast (*18*, *19*). However, in those assays, an unprocessed D-loop could merge with a converging replication fork to form a single Holiday junction (HJ), which cannot be removed by Sgs1-Top3-Rmi1, likely requiring resolution by Mus81 or Yen1. In our assays with fork breakage at subtelomeric positions (ChrII and ChrIV), where only inefficient or dormant origins are present (fig. S1), the unprocessed D-loop can migrate to the end of the chromosome. In this context, *mus81*Δ *yen1*Δ cells are more resistant to fork breakage, and this improvement depends on Pol32 (Fig. 1B and fig. S3). A likely explanation for the mild repair defect in *mus81*Δ *yen1*Δ cells at subtelomeric positions, with Pol32-BIR remaining insufficient for repair, is the firing of normally inefficient or dormant origins (*20*, *21*), resulting in the merging of a converging fork with the extending D-loop. Last, key recombination proteins Rad52 and Rad51 are needed for broken fork repair (Fig. 1B and fig. S4). Together, the genetic requirements suggest that Pol32-mediated BIR is neither an efficient nor the primary pathway of broken fork repair and is suppressed by converging forks.

The other defining characteristic of BIR is a high level of template switches and point mutations at least 10 or 30 kb from the strand invasion site, respectively (*2*, *3*). We therefore determined whether broken fork repair is marked by the same events using *URA3* reporters at different distances from the FRT sites (Fig. 2A). *URA3* disruption confers resistance to 5-fluoroorotic acid (5-FOA); therefore, we determined the rate of 5-FOA^R–conferring mutations after repair. Two versions of these assays were made, either with or without the *ura3-1* point mutant allele at its natural locus on ChrV, which allows for observation of recombination template switches. Conversion to *ura3-1* requires initial DNA synthesis at the sister chromatid template to reach *URA3*, followed by a switch to *ura3-1* and either mismatch correction or DNA synthesis copying the base pair change. To recover these events, a second switch back to the original sister chromatid is required because single-switch products followed by DNA synthesis are inviable for nearly all reporter positions. The exception is the position 15 kb from the FRT on ChrII, where a single template switch to *ura3-1* followed by synthesis to the end of ChrV yields a viable repair product (Fig. 2B).

In wild-type cells, mutation rates are significantly increased next to the broken fork but drop to spontaneous levels at distances beyond 10 kb from the DSB. The same pattern is observed for all break locations (Fig. 2C), indicating a shift in the fidelity of synthesis during repair. When the *ura3-1* allele is present, mutation rates are further elevated, but these increases remain in proximity to the break (Fig. 2D).

To characterize mutations, we sequenced reporters positioned 0.2 or 15 kb from the broken fork on ChrII from 40 to 60 independent 5-FOA^R colonies. Three primary categories of variants were scored: point mutations, conversions to

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [2]Department of Molecular Biology and Genetics, University of Aarhus, Aarhus 8000, Denmark. [3]Department of Pediatrics, and Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [4]Texas Children's Hospital, Houston, TX 77030, USA. *Present address: Graduate School of Biomedical Sciences at Houston, University of Texas, 6767 Bertner Avenue, Houston, TX 77030, USA. †Corresponding author. E-mail: gira@bcm.edu

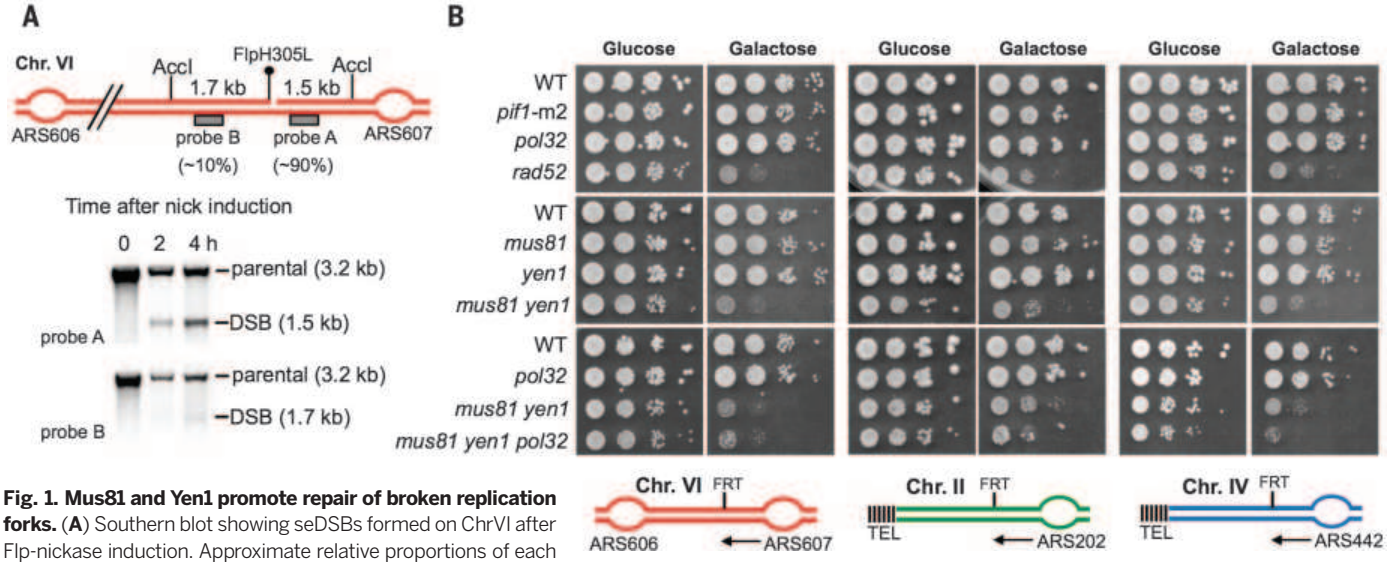


Fig. 1. Mus81 and Yen1 promote repair of broken replication forks. (**A**) Southern blot showing seDSBs formed on ChrVI after Flp-nickase induction. Approximate relative proportions of each DSB band are indicated below the location of the corresponding probe. (**B**) Serial 10-fold dilutions plated on YP-glucose and YP-galactose (Flp induction) to assay for repair defects in mutant strains. Arrows indicate the predominant direction of replication forks reaching the FRT.

ura3-1, and gross chromosomal rearrangements/deletions/duplications (called hereafter GCRs), where the reporter gene could not be amplified or had a different size. We also analyzed spontaneous mutagenesis using cells without the Flp recombinase. We observed a 36- to 53-fold increase in point mutations and a 186-fold increase in *ura3-1* conversions next to the break but no increase 15 kb from the broken fork (Fig. 2E). Frameshift mutations are increased threefold more (95×) than base substitutions (35×) (table S1), which is indicative of compromised mismatch repair (*22*). GCRs are increased 100-fold without *ura3-1* and an additional 100-fold in the presence of *ura3-1*, indicating that *ura3-1* is used as a recombination template to generate GCRs (Fig. 2E).

To establish whether cleavage of recombination intermediates is required for the fidelity of repair, we measured mutation rates in *mus81*Δ and *yen1*Δ mutants. In both, mutation rates remain similar to that of wild-type close to the fork breakage but increase significantly, by 6- to 95-fold, in *mus81*Δ cells farther from the broken fork at subtelomeric locations (Fig. 3B). In the presence

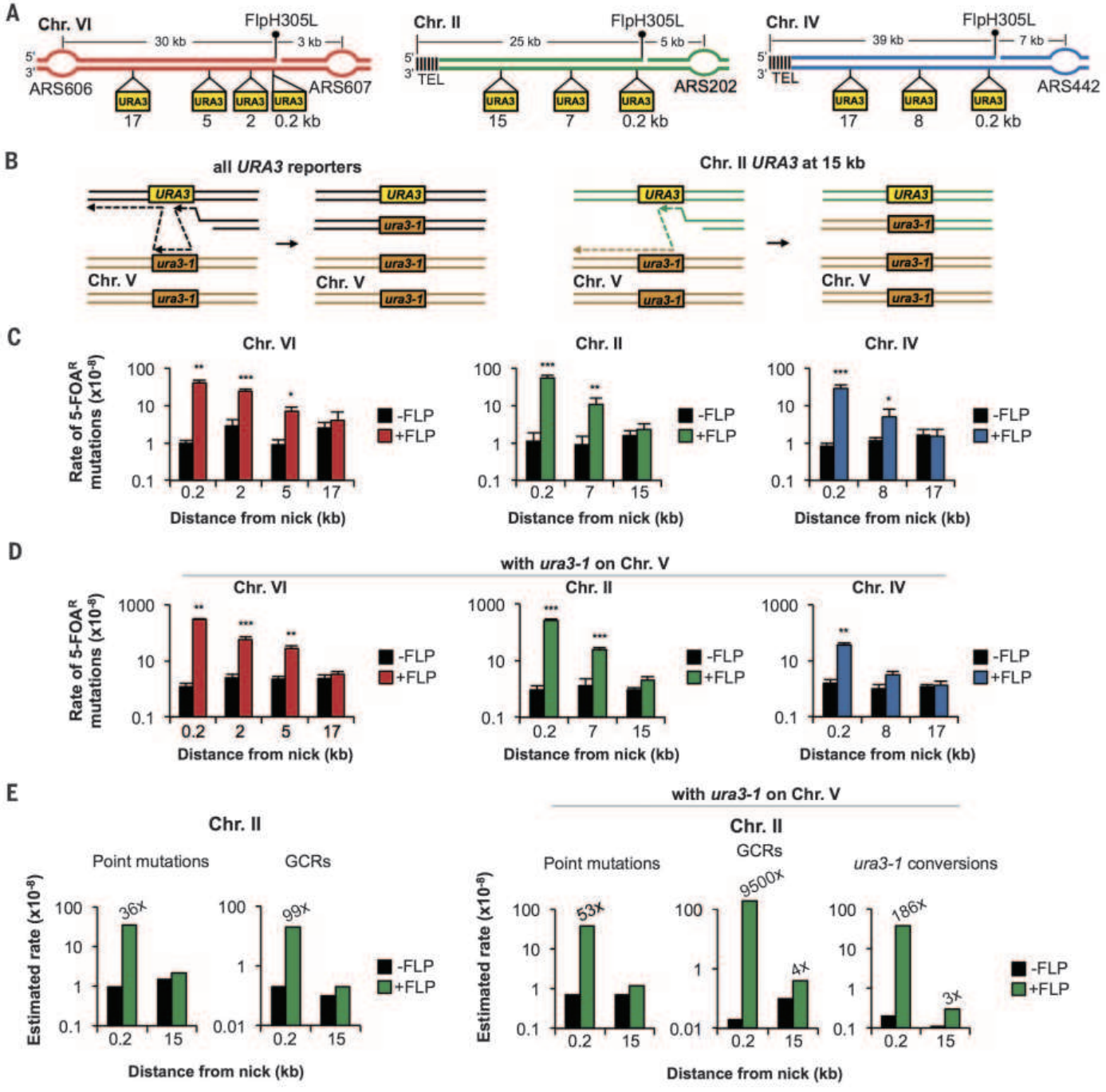


Fig. 2. Mutagenesis in wild-type cells is restricted to the vicinity of the broken replication fork. (**A**) Positions of *URA3* reporter genes with respect to broken forks. (**B**) Diagram of possible recombination template switches leading to gene conversion. (**C** and **D**) Median 5-FOA^R break–induced (+FLP) and spontaneous (−FLP) mutation rates in wild-type cells (C) without or (D) with *ura3-1* on ChrV. Error bars show 1 SD, and significant increases over spontaneous are indicated. *P < 0.05, **P < 0.01, ***P < 0.001. (**E**) Estimated rates of point mutations, GCRs, and *ura3-1* conversions. Fold changes above spontaneous are indicated.

of the *ura3-1* allele, the increase of mutations is dramatic even close to the break in *mus81*Δ cells, regardless of FRT positions. Farther from the break, an increase of up to 150-fold is observed, but only at subtelomeric positions (Fig. 3C). These results are consistent with a major role of Mus81 in limiting mutagenic repair, which is particularly important when there is no efficient converging fork.

Sequencing of reporters from 5-FOA^R cells at positions next to broken forks (ChrII, ChrVI, and ChrIV) or 15 to 17 kb away (ChrII and ChrVI) revealed that the most prominent change in the absence of Mus81 is an increased usage of the *ura3-1* allele during repair. Resultant conversion events increase fivefold over that of wild type close to the break, with an efficient converging fork. At subtelomeric positions, conversions increase 18- or 96-fold near the break and up to 115-fold 15 kb from the break (Fig. 4, A and B, and fig. S5). Neither GCRs nor point mutations are increased more than twofold in *mus81*Δ cells next to the break site. In contrast, both GCRs and point mutations are increased above wild-type levels farther from the break at subtelomeric positions (Fig. 4, A and B, and fig. S5). Thus, an efficient converging fork dictates the distance of inaccurate synthesis in *mus81*Δ cells. Most GCRs (10 out of 10 tested) in *mus81*Δ cells at the *URA3* reporter 15 kb from the broken fork on ChrII result from nonreciprocal translocations with *ura3-1*, as shown through analysis of rearrangement junctions (fig. S6), further documenting switches to nonallelic templates in *mus81*Δ cells, even far from the break.

Although *yen1*Δ cells have no notable change in any type of event, elimination of Yen1 in a *mus81*Δ background, tested only at ChrII, leads to a further increase of GCRs but no increase in template switches and only a mild increase in point mutations (Fig. 4B and fig. S5). These results support the view that Yen1 plays a backup

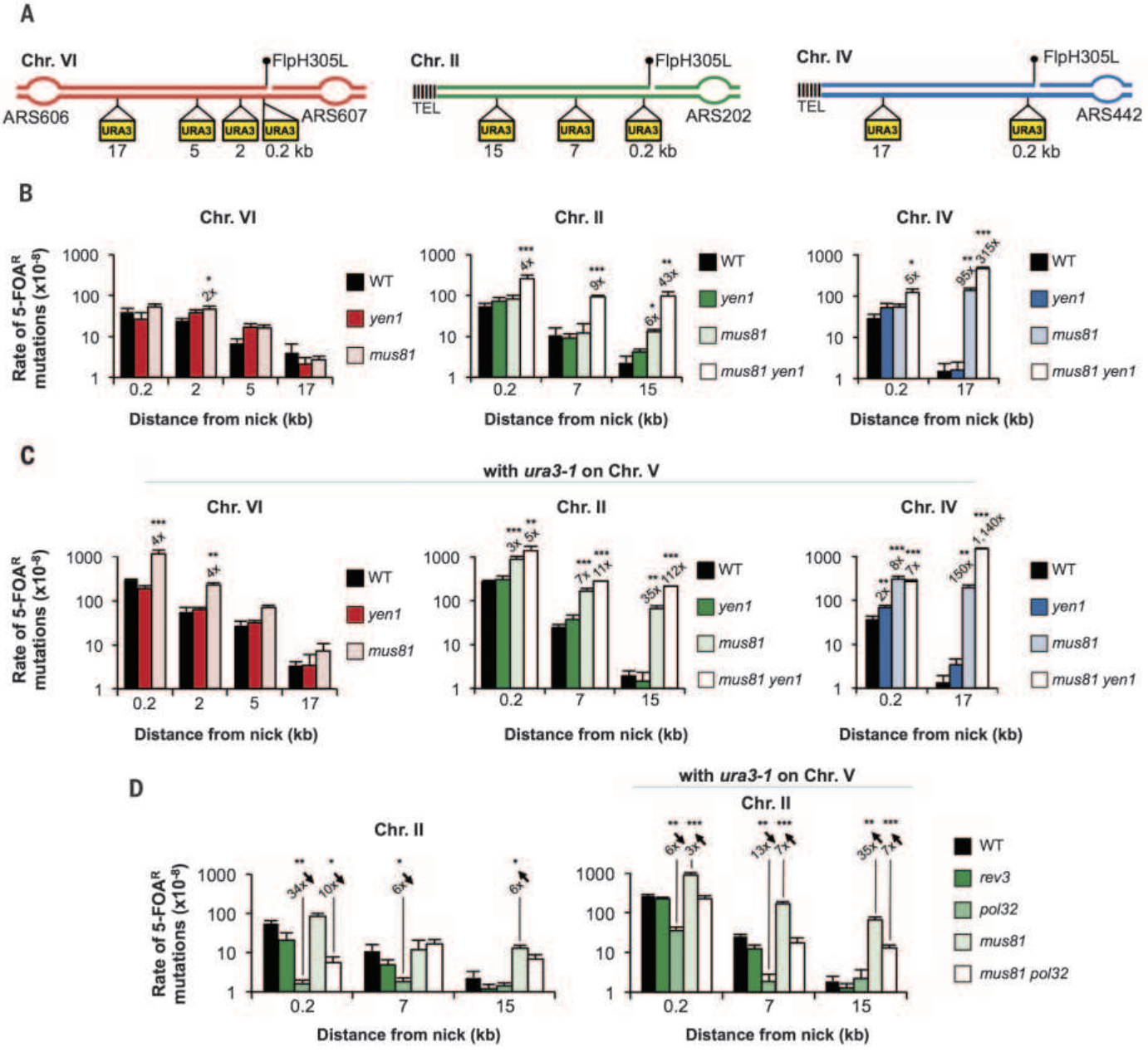


Fig. 3. Mus81 suppresses error-prone Pol32-dependent synthesis at broken replication fork. (**A**) Location of *URA3* reporters used for mutation analysis. (**B** to **D**) Median 5-FOA^R mutation rates in wild-type (WT) cells or indicated mutants. Significant fold changes compared with wild type are indicated, and error bars represent 1 SD. *P < 0.05, **P < 0.01, ***P < 0.001.

role in the absence of Mus81 (*23*), likely processing structures that arise in the absence of Mus81, which are outcomes of the D-loop merging with a converging fork. This is consistent with the late M-phase activation of Yen1 (*24*, *25*).

BIR is mediated by Rad51 and Pol32, and indeed the majority of mutations at broken forks in wild-type and *mus81*Δ cells arise in a Rad51- and Pol32-dependent manner (Fig. 3D and fig. S4). This indicates that Pol32 mediates initial DNA synthesis at broken forks and that Mus81 decreases the distance it travels. Consistent with high Pol32-dependent mutation rates in *mus81*Δ, Pol32 plays a role in resistance to single-fork breakage induced at subtelomeric locations or by high doses of CPT (fig. S2), suggesting that repair proceeds via Pol32-BIR in the absence of Mus81 and an efficient converging fork. Pol32 also works with Polζ, encoded by *REV3*, but this function of Pol32 is not important at broken forks, as *rev3*Δ does not show a significant decrease in mutations (Fig. 3D and fig. S7). Sequencing analysis revealed that nearly all break-induced point mutations in wild-type and *mus81*Δ cells depend on Pol32. Also, most events involving recombination template switches to *ura3-1* depend on Pol32, although some GCRs clearly are Pol32-independent (fig. S7).

The presence of *ura3-1* greatly increases GCRs next to the FRT on both ChrII and ChrVI in wild-type cells. Rearrangements in both cases involve deletion of sequences between the *URA3* reporter and the closest Ty long terminal repeat (LTR) on the opposite side of the FRT. The deleted regions are replaced by sequences between *ura3-1* and a Ty LTR located on ChrV (figs. S8 and S9). Because these HR events use substrates on both sides of the FRT and, unlike recombination template switches, are not substantially affected by *MUS81* deletion, it is likely that they correspond to repair of rare two-ended breaks by Pol32-dependent gap repair (*26*).

More than half of the human genome is composed of interspersed repeats, with *Alu* repetitive elements being the most abundant class (~1.1 million copies). *Alu*s are often found at breakpoints of copy number variants as well as complex genomic rearrangements associated with disease (*27*). To address whether broken fork repair stimulates template switches between diverged *Alu* elements, we inserted two *Alu*s (88.2% identical) that were previously observed to mediate human disease-associated deletions (*28*) onto ChrII, 1.2 and 9.7 kb from the FRT site, flanking a *URA3* reporter (Fig. 4C). *Alu-Alu* template switches lead to deletion of the 8.5-kb sequence separating them. In wild-type cells, induction of broken forks increases 5-FOA^R mutation rates by ~10-fold, the majority of which are *Alu-Alu* template switches. Deletion of *MUS81* further increases the rate of *Alu-Alu* template switches 83-fold above that of wild type (Fig. 4D), underscoring the role of Mus81 in suppressing switches between highly diverged templates. Switches between *Alu*s occurred at 4 to 21 nucleotides of microhomology (fig. S10). These results suggest that template switch mechanisms could account for a substantial proportion of *Alu-Alu* structural variation in humans, rather than unequal crossing over.

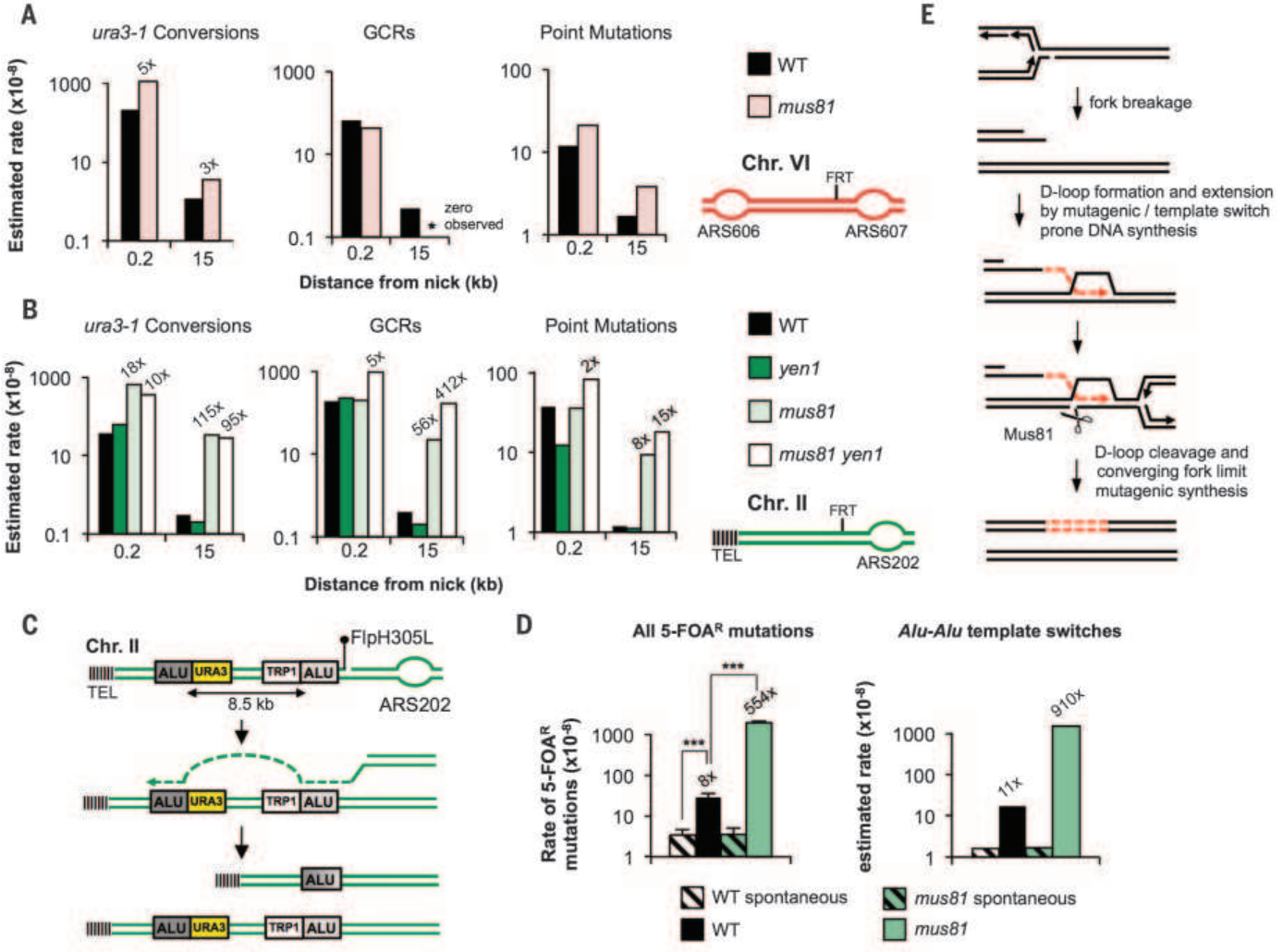


Fig. 4. Mus81 prevents excessive donor site–switching during repair. (**A** and **B**) Estimated rates of mutation categories at broken replication forks on (A) ChrVI and (B) ChrII. Fold changes over wild type are indicated. (**C**) Scheme of *Alu*-mediated template switches resulting in deletion. (**D**) Total 5-FOA^R rates and estimated rates of *Alu*-mediated template switches. (**E**) Model for the control of repair fidelity at broken replication forks.

We demonstrate that fidelity of repair at broken replication forks depends on two partially compensatory mechanisms: cleavage by Mus81 and arrival of a converging fork (Fig. 4E and fig. S11). Converging forks limit the need to reestablish fully functional forks, illustrating an advantage of the multi-origin nature of eukaryotic chromosomes. We propose that deficiencies in Mus81 or timely converging forks may underlie the increased usage of POLD3/Pol32-mediated BIR in cancer cells (*9*) and consequently provide higher adaptation potential to cancer cells and promote tumor progression.

REFERENCES AND NOTES

1. C. Shee, J. L. Gibson, S. M. Rosenberg, *Cell Rep.* **2**, 714–721 (2012).
2. A. Deem *et al.*, *PLOS Biol.* **9**, e1000594 (2011).
3. C. E. Smith, B. Llorente, L. S. Symington, *Nature* **447**, 102–105 (2007).
4. F. Pratto *et al.*, *Science* **346**, 1256442 (2014).
5. J. R. Lydeard, S. Jain, M. Yamaguchi, J. E. Haber, *Nature* **448**, 820–823 (2007).
6. N. Saini *et al.*, *Nature* **502**, 389–392 (2013).
7. M. A. Wilson *et al.*, *Nature* **502**, 393–396 (2013).
8. C. J. Sakofsky *et al.*, *Cell Rep.* **7**, 1640–1648 (2014).
9. L. Costantino *et al.*, *Science* **343**, 88–91 (2014).
10. C. M. Carvalho *et al.*, *Nat. Genet.* **45**, 1319–1326 (2013).
11. I. Nielsen *et al.*, *Nat. Methods* **6**, 753–757 (2009).
12. S. R. McGuffee, D. J. Smith, I. Whitehouse, *Mol. Cell* **50**, 123–135 (2013).
13. A. R. Clausen *et al.*, *Nat. Struct. Mol. Biol.* **22**, 185–191 (2015).
14. R. A. Donnianni, L. S. Symington, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13475–13480 (2013).
15. K. T. Ehmsen, W. D. Heyer, *Nucleic Acids Res.* **36**, 2182–2195 (2008).
16. A. Pepe, S. C. West, *Cell Rep.* **7**, 1048–1055 (2014).
17. C. K. Ho, G. Mazón, A. F. Lam, L. S. Symington, *Mol. Cell* **40**, 988–1000 (2010).
18. S. Muñoz-Galván *et al.*, *Mol. Cell. Biol.* **32**, 1592–1603 (2012).
19. L. Roseaulin *et al.*, *EMBO J.* **27**, 1378–1387 (2008).
20. Y. Doksani, R. Bermejo, S. Fiorani, J. E. Haber, M. Foiani, *Cell* **137**, 247–258 (2009).
21. C. A. Müller *et al.*, *Nucleic Acids Res.* **42**, e3 (2014).
22. A. Malkova, J. E. Haber, *Annu. Rev. Genet.* **46**, 455–473 (2012).
23. M. G. Blanco, J. Matos, U. Rass, S. C. Ip, S. C. West, *DNA Repair (Amst.)* **9**, 394–402 (2010).
24. M. G. Blanco, J. Matos, S. C. West, *Mol. Cell* **54**, 94–106 (2014).
25. C. L. Eissler *et al.*, *Mol. Cell* **54**, 80–93 (2014).
26. S. Jain *et al.*, *Genes Dev.* **23**, 291–303 (2009).
27. S. Gu *et al.*, *Hum. Mol. Genet.* **24**, 4061–4077 (2015).
28. P. M. Boone *et al.*, *Am. J. Hum. Genet.* **95**, 143–161 (2014).

ACKNOWLEDGMENTS

We thank P. Hastings, A. Malkova, S. Rosenberg, and D. Bates for critical comments on the manuscript. This work was supported by National Institutes of Health (NIH) grants GM080600 to G.I., NS058529 and HG006542 to J.R.L., and NS083159 to I.M.C. C.R.B is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation (DRG 2155-13). J.R.L. is a paid consultant for Regeneron Pharmaceuticals and has stock options in Lasergen.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/742/suppl/DC1
Materials and Methods
Figs. S1 to S11
Table S1
References (*29*–*31*)

3 February 2015; accepted 14 July 2015
10.1126/science.aaa8391

EVOLUTION

Fruit flies diversify their offspring in response to parasite infection

Nadia D. Singh,[1]* Dallas R. Criscoe,[2] Shelly Skolfield,[3] Kathryn P. Kohl,[4] Erin S. Keebaugh,[5] Todd A. Schlenke[3]*

The evolution of sexual reproduction is often explained by Red Queen dynamics: Organisms must continually evolve to maintain fitness relative to interacting organisms, such as parasites. Recombination accompanies sexual reproduction and helps diversify an organism's offspring, so that parasites cannot exploit static host genotypes. Here we show that *Drosophila melanogaster* plastically increases the production of recombinant offspring after infection. The response is consistent across genetic backgrounds, developmental stages, and parasite types but is not induced after sterile wounding. Furthermore, the response appears to be driven by transmission distortion rather than increased recombination. Our study extends the Red Queen model to include the increased production of recombinant offspring and uncovers a remarkable ability of hosts to actively distort their recombination fraction in rapid response to environmental cues.

The first observation that the proportion of recombinant offspring produced by individuals could vary in response to environmental conditions was made in *Drosophila* nearly 100 years ago (*1*). Evidence continues to accumulate that recombination frequency in a variety of species plastically varies in response to factors such as maternal age, temperature, nutritional status, and social stress (*2*–*4*). Theoretical models indicate that plastic recombination can evolve if organismal fitness and recombination frequency are negatively correlated (*5*). Such a negative correlation enables maintaining beneficial combinations of alleles on linked haplotypes while providing opportunities for less fit combinations of alleles to be disrupted and reassembled into potentially more fit haplotypes. Although this fitness-associated recombination model (*5*) appears theoretically tractable for haploids, the model is less applicable for diploids (*6*), because haplotype fitness and organismal fitness are not equivalent. Thus, in spite of numerous observed instances of plastic recombination, a general explanation for its origin and maintenance in natural populations remains elusive.

The evolutionary advantage of sexual reproduction itself, and the independent chromosome segregation and recombination that accompany sex, remain a hotly debated topic in biology. Why cede half of your genetic inheritance to a partner when producing offspring, and shuffle beneficial combinations of alleles that allowed you to survive and reproduce? A leading hypothesis for the evolution of sex is the Red Queen hypothesis, which argues that sex is favored in the face of dynamic selection pressures (*7*, *8*), such as antagonistic interactions with coevolving organisms. Sex and recombination allow parents to diversify their offspring so that competitors, predators, and parasites cannot exploit a static competitor/prey/host genotype. Parasites in particular are thought to be especially important in this process, given that they usually have shorter generation times than their hosts and so can evolve more rapidly. There is strong evidence that host species experiencing parasite-mediated selection pressures are more likely to evolve sexuality (*9*), increased outcrossing (*10*), and increased recombination rates (*11*). There is less evidence that host individuals, during their lifetimes, plastically increase their sexual reproduction (*12*), outcrossing (*13*), or recombination rates in response to parasite threats. In plants, somatic recombination frequency plastically increases in response to immune stress (*14*), and infection leads to a variety of meiotic perturbations, including increased meiotic recombination frequency (*15*). However, there is no direct evidence to date for parasitism-induced plastic increases in meiotic recombination frequency in animals. We used the fruit fly *D. melanogaster* to test whether the proportion of recombinant offspring plastically increases after parasite infection, as predicted by the Red Queen model. We explored parasite-associated plastic recombination using different types of parasites and stresses, in different host life stages, and using different host genetic backgrounds. Further, we began to dissect the mechanism behind the parasite-mediated increase in recombination fraction that we observed.

Borrowing from the rich history of recombination frequency estimation in *D. melanogaster*, we tested whether infection plastically increases the recombination fraction using a classical genetic approach (Fig. 1). We crossed a wild-type strain to a strain with recessive, visible mutations

[1]Department of Biological Sciences and Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA. [2]Translational Biology and Molecular Medicine Program, Baylor College of Medicine, Houston, TX, USA. [3]Department of Biology, Reed College, Portland, OR, USA. [4]Department of Biology, Winthrop University, Rock Hill, SC, USA. [5]Department of Biology, Emory University, Atlanta, GA, USA.
***Corresponding author. E-mail: ndsingh@ncsu.edu (N.D.S.); schlenkt@reed.edu (T.A.S.)**

in the genes *ebony* and *rough*, which reside approximately 20 centimorgans (cM) apart on chromosome 3R. Female F_1 progeny from this cross are doubly heterozygous, and gametic recombination events occurring between *ebony* and *rough* in these F_1 females were scored after treatment by examining their offspring. To reveal recombination events, the F_1 females were backcrossed to double-mutant males, and backcross 1 individuals were scored: Those carrying a maternal chromosome with a single crossover event between the two markers will have one visible mutation but not the other (Fig. 1).

We first infected virgin F_1 adult females by piercing them in the thorax with a needle dipped in a culture of *Serratia marcescens,* a Gram-negative bacterium that opportunistically infects a wide range of hosts (*16*). We mock-infected a control group with sterile media. The experiment was conducted with four randomly selected wild-type strains from the *Drosophila melanogaster* Genetic Reference Panel (*17*). Four days after treatment, these F_1 females were individually paired with double-mutant males and allowed to mate and oviposit for 5 days. Between 8 and 16 replicate F_1 females were used per treatment per line, which yielded a total of 14,732 progeny scored in the experiment (table S1). The genotype, treatment, and genotype-by-treatment interaction effects on the proportion of recombinant offspring were evaluated using a generalized linear model (*18*). Consistent with the Red Queen hypothesis, treatment explained a significant proportion of the observed variance in recombination fraction ($P = 0.03$, χ^2 test, Fig. 2), with infected flies producing a higher proportion of recombinant offspring than wounded controls. This trend was consistent across all four genetic backgrounds (mean recombination fractions were 0.200 versus 0.178, 0.180 versus 0.178, 0.191 versus 0.187, and 0.199 versus 0.179 for infected versus control flies in lines RAL21, RAL40, RAL45, and RAL75, respectively). No other factors or interactions significantly contributed to the observed variation in recombination fraction in this experiment ($P > 0.35$, all factors, χ^2 test).

The post-infection increase in recombinant offspring we observed may be caused by an increase in meiotic recombination, by an increase in germ cell mitotic recombination, or by some form of transmission distortion, whereby recombinant gametes and/or embryos outnumber or outcompete nonrecombinants (via viability differences among genotypes or asymmetric meiosis where nonrecombinant chromosomes preferentially end up in polar bodies). To test whether the plastic increase in recombination fraction was due to increased mitotic recombination, we conducted a second experiment exploiting the fact that *Drosophila* males do not undergo meiotic recombination (*19*). We infected doubly heterozygous F_1 males with *S. marcescens* and, immediately after treatment, backcrossed these males to double-mutant females. Out of the 9218 flies scored in this experiment, we observed no recombinant progeny. This indicates that there is an overall very low mitotic recombination rate in the *Drosophila* germ line and that bacterial infection does not detectably elevate the rate, at least in males. These data suggest that increased meiotic recombination or transmission distortion, not increased mitotic recombination, are preferred explanations for the increased recombination fraction observed in infected females.

To determine whether transmission distortion contributed to our observations, we relied on the known timing of recombination during *Drosophila* oogenesis: *D. melanogaster* crossovers are initiated and resolved in developing oocytes 4 to 5 days before eggs are fertilized and laid (*20*, *21*). Thus, any increase in the recombination fraction that manifests in the first 4 days after infection must primarily be driven by transmission distortion rather than increased crossing-over during meiotic prophase. We conducted an experiment in which heterozygous females were infected 48 hours after mating and serially transferred to new vials every 4 days for 12 days. Two additional treatments were included in this experiment: a no-treatment control and an infection treatment using a second bacterial species, *Providencia rettgeri*, a Gram-negative bacterium that infects *D. melanogaster* in nature (*22*). Between 120 and 183 replicate females were used per treatment, yielding a total of 32,256 progeny scored in this experiment (table S1). Because these data were significantly overdispersed ($P < 0.001$, Pearson goodness of fit), we used a generalized linear model that includes an overdispersion parameter to test the effects of treatment and time on the mean recombination fraction (*18*). We found a significant effect of treatment ($P = 0.02$, χ^2 test). There was also a significant effect of time ($P = 0.01$, χ^2 test), as expected given that the recombination fraction shows maternal age effects in *Drosophila* (*2*). The interaction between treatment and time was not significant ($P = 0.44$, χ^2 test).

Bolstered by the significant differences in the recombination fraction across treatments, we sought to test predictions of the Red Queen hypothesis using comparisons of means between specific treatment pairs. Consistent with the Red Queen hypothesis, females infected with *S. marcescens* had a significantly higher recombination fraction relative to the sterile wound control in both the 1- to 4-day (mean recombination fractions of 0.211 versus 0.178) and 5- to 12-day (0.195 versus 0.165) collection periods ($P = 0.03$, $P = 0.04$, respectively, two-tailed t test) (Fig. 3). Females infected with *P. rettgeri* showed similar results, with a marginally significant increase in the recombination fraction relative to the sterile wound control in the 1- to 4-day collection period (0.204 versus 0.178; $P = 0.06$, two-tailed t test) and a significant increase in the 5- to 12-day collection period (0.195 versus 0.165; $P = 0.04$, two-tailed t test). The no-treatment control was uninformative in both collection periods, showing no significant difference in recombination fraction relative to the sterile wound or either bacterial treatment ($P > 0.24$, all comparisons). Nonparametric comparisons of means echo these findings (table S2), highlighting the robustness of our results to assumptions regarding the distribution of the error terms. Overall, these data confirm our initial findings of an infection-associated increase in recombination fraction and extend them to a new bacterial parasite. Because this effect manifests in the first 4 days after infection, it is unlikely that the increase in recombination fraction is caused by an increase in the frequency of crossing-over during meiosis; instead, it is probably due to transmission distortion. A rapid increase in the frequency of recombinant progeny consistent with transmission distortion was also observed in *D. melanogaster* in response to heat shock and multiple mating (*23*, *24*). Our data reveal a remarkable ability of hosts to alter their recombination fraction in rapid response to environmental cues.

To determine whether infection by a different kind of parasite, which infects a different host life stage, can also induce a plastic increase in host recombination frequency, we exposed F_1 doubly heterozygous female (Fig. 1) larvae to the parasitic wasp *Leptopilina clavipes*. This wasp lays a single egg in the body cavity of larval flies, which then hatches and consumes the fly from the inside out unless it is melanotically encapsulated and killed by host hemocytes (*25*). Fly larvae that successfully fought off wasp infection were identified in the adult stage by the presence of black capsules in their abdomens. Virgin wasp-infected and control F_1 females were backcrossed to doubly

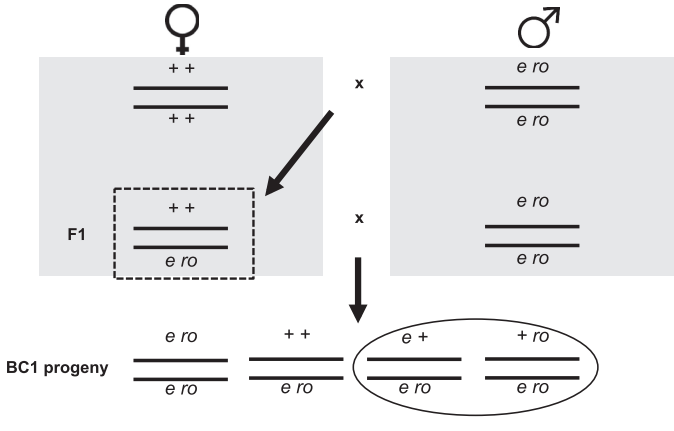


Fig. 1. Schematic representation of the two-step crossing scheme using *ebony* (e) and *rough* (*ro*). Females used in each cross are shown on the left, males on the right. F_1 doubly heterozygous females (dashed box) are those that were subjected to parasite or control treatments. Backcross 1 (BC1) progeny with either of the two recombinant genotypes (circled) can be visually identified using our screen.

marked males (Fig. 1) and were serially transferred into oviposition vials in groups of three individuals every 2 days for 12 days. A total of 87 control and 69 wasp-infection replicates were used in this experiment, which yielded a total of 50,140 progeny (table S1).

We used a generalized linear model to test the effects of treatment and time on the mean recombination fraction. We found a significant effect of treatment (P = 0.0002, χ^2 test), with wasp-infected flies producing a greater proportion of recombinant offspring (0.238) than uninfected controls (0.222). There was no significant effect of time or the interaction between treatment and time (P = 0.15, P = 0.38, respectively, χ^2 test). When analyzed separately for each 2-day egg-laying period, the recombination fraction was significantly increased in wasp-infected flies for the post-mating time periods spanning days 3 to 4 and 5 to 6 (Fig. 4B) (P = 0.02, both comparisons, two-tailed t test). Nonparametric comparisons of means verified these findings (table S2). These data indicate that wasp infection of larval flies, like bacterial infection of adults, leads to a plastic increase in recombinant offspring, once again consistent with the Red Queen model.

This result is surprising given that *Drosophila* larvae contain only primordial ovaries. Ovarioles in *D. melanogaster* females do not begin to develop until after pupariation (*26*); differentiation of the germarial regions (in which crossing-over occurs) within each ovariole takes place in the 24 hours after puparium formation, and synaptonemal complexes (structures required for crossing-over in wild-type *Drosophila*) in the first pro-oocytes become visible in the developing ovaries at 36 hours after puparium formation (*26*). The marked delay between wasp infection and the onset of oocyte formation, coupled with the observation that the recombinational response to the wasp attack is sustained for up to 12 days after mating, shows that the signal underlying the infection-associated increase in recombination fraction can be triggered in the absence of fully developed ovaries and can be sustained across development.

Our data indicate that the proportion of recombinant offspring in *D. melanogaster* plastically increases in response to a variety of parasite pressures. The onset of the response can be both rapid and prolonged, as infected adults increase their recombination fraction within 1 to 4 days after bacterial infection, and wasp-infected larvae develop for several days and undergo metamorphosis before they begin laying eggs at all. Data from both bacteria and wasp-infection trials show that the increased production of recombinant offspring lasts for several days and, for bacterial infection, is significantly stronger than any effect induced by a sterile wound. Because female flies exposed to heat-killed bacteria only show a weak, nonsignificant elevation in recombination fraction relative to wounded flies (0.189 versus 0.178; P = 0.28, two-tailed t test, fig. S1 and table S2), active parasite signals or host immune signaling pathways that specifically respond to live parasites appear to be required for the full-blown recombination response. Furthermore, the

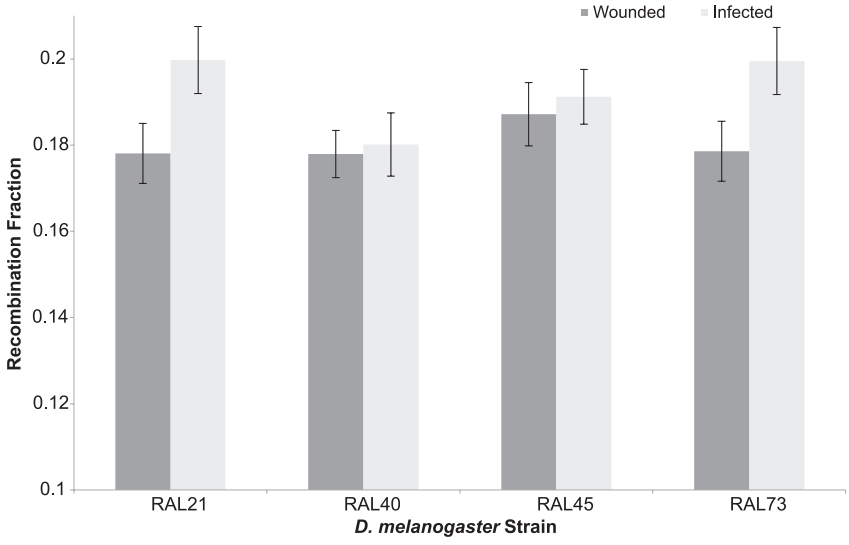


Fig. 2. Recombination fraction for four wild-type strains of *D. melanogaster* mock-infected (wounded) or infected with *S. marcescens*. Error bars represent standard error.

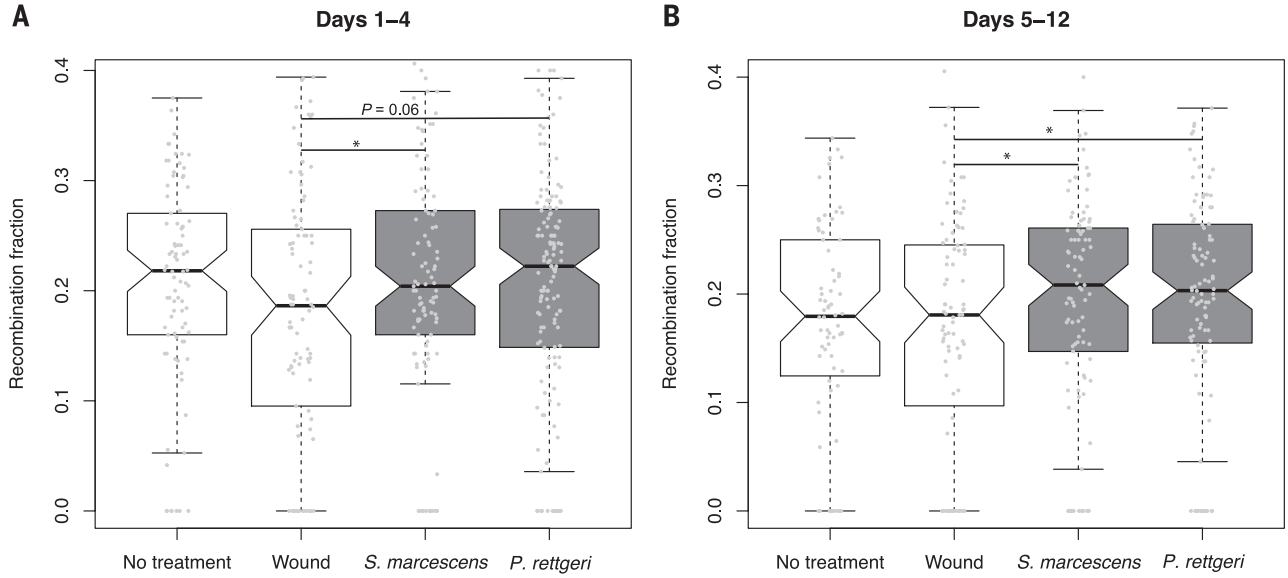


Fig. 3. Box plots illustrating the distribution of recombination fractions in *D. melanogaster* strain RAL73 after one of four treatments: no treatment, sterile wound, infection with *S. marcescens*, or infection with *P. rettgeri*. The median is marked with a black line; the first and third quartiles are represented as the lower and upper edges of the box, respectively. The whiskers extend to the most extreme data point no farther from the box than 1 times the interquartile range. Jittered, individual data points are presented as gray circles. Recombination was estimated separately for eggs laid (**A**) days 1 to 4 after infection and (**B**) days 5 to 12 after infection. Pairwise comparisons of transformed data that are statistically significant based on a two-tailed t test at $P \leq 0.05$ are marked with an asterisk.

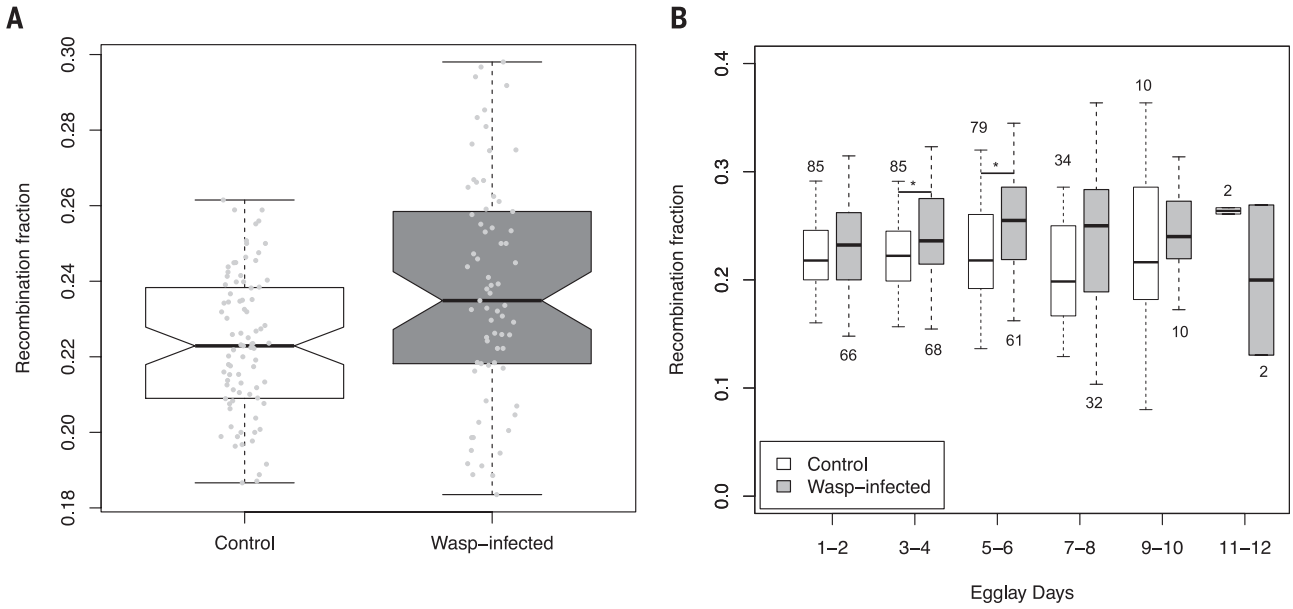


Fig. 4. Box plots illustrating the distribution of recombination fractions in *D. melanogaster* strain RAL73 in control and wasp-infected females. The median is marked with a black line; the first and third quartiles are represented as lower and upper edges of the box, respectively. The whiskers extend to the most extreme data point no farther from the box than 1 times the interquartile range. Recombination fraction is shown (**A**) estimated over the entire 12-day egg-laying period and (**B**) in each of the six 2-day egg-laying periods. In (A), jittered, individual data points are presented as gray circles. In (B), the number of replicates for each time point is included for the control (above the top whisker) and the wasp-infected (below the bottom whisker) treatments. Because there are only two replicates for the 11- to 12-day period, the edges of the box completely span the range of observations. Pairwise comparisons of transformed data that are statistically significant based on a two-tailed *t* test at $P \leq 0.05$ are marked with an asterisk.

increase in recombination fraction is not driven by viability defects caused by an interaction between infection status and the visible markers used in this study (supplementary materials). We find that the mechanism underlying the increase in recombinant offspring is transmission distortion. This distortion could be due to asymmetries during meiosis II or to viability differences between recombinant and nonrecombinant gametes or progeny, and represents an as yet unappreciated mechanism by which *D. melanogaster* females plastically alter the frequency of the recombinant progeny they produce. In the future, it will be important to identify the mechanisms by which this distortion is mediated, as well as determine the extent to which the plastic increase in recombination fraction observed in the current study extends genome-wide, given that previous work has shown that stress-induced changes in recombination frequency are not uniform across the genome (*2*). Overall, our work identifies a strong link between infection and recombination in animals and further extends the Red Queen hypothesis to include plastic changes in recombination in response to environmental stimuli.

REFERENCES AND NOTES

1. C. B. Bridges, *J. Exp. Zool.* **19**, 1–21 (1915).
2. C. Stern, *Proc. Natl. Acad. Sci. U.S.A.* **12**, 530–532 (1926).
3. J. V. Neel, *Genetics* **26**, 506–516 (1941).
4. D. K. Belyaev, P. M. Borodin, *Biol. Zent. Bl.* **101**, 705 (1982).
5. L. Hadany, T. Beker, *Genetics* **165**, 2167–2179 (2003).
6. A. F. Agrawal, L. Hadany, S. P. Otto, *Genetics* **171**, 803–812 (2005).
7. J. Jaenike, *Evol. Theory* **3**, 191 (1978).
8. W. D. Hamilton, *Oikos* **35**, 282 (1980).
9. C. M. Lively, *Nature* **328**, 519–521 (1987).
10. J. W. Busch, M. Neiman, J. M. Koslow, *Evolution* **58**, 2584–2590 (2004).
11. M. Greeff, P. Schmid-Hempel, *Genetica* **138**, 737–744 (2010).
12. R. Mostowy, J. Engelstädter, *J. Evol. Biol.* **25**, 2033–2046 (2012).
13. D. M. Soper, K. C. King, D. Vergara, C. M. Lively, *Biol. Lett.* **10**, 20131091 (2014).
14. I. Kovalchuk *et al.*, *Nature* **423**, 760–762 (2003).
15. L. Andronic, *Can. J. Plant Sci.* **92**, 1083–1091 (2012).
16. P. A. D. Grimont, F. Grimont, *Annu. Rev. Microbiol.* **32**, 221–248 (1978).
17. T. F. C. Mackay *et al.*, *Nature* **482**, 173–178 (2012).
18. Materials and methods are available as supplementary materials on *Science* Online.
19. T. H. Morgan, *Biol. Bull.* **26**, 195 (1914).
20. S. Mehrotra, R. S. Hawley, K. S. McKim, in *Genome Dynamics and Stability*, R. Egel, D. H. Lankenau, Eds. (Springer-Verlag Berlin, Berlin, Germany, 2008), vol. 2, pp. 125–151.
21. R. Bhagat, E. A. Manheim, D. E. Sherizen, K. S. McKim, *Cytogenet. Genome Res.* **107**, 160–171 (2004).
22. K. A. McKean, C. P. Yourth, B. P. Lazzaro, A. G. Clark, *BMC Evol. Biol.* **8**, 76 (2008).
23. N. K. Priest, D. A. Roach, L. F. Galloway, *Evolution* **61**, 160–167 (2007).
24. W. H. Zhong, N. K. Priest, *Behav. Ecol. Sociobiol.* **65**, 493–502 (2011).
25. Y. Carton, M. Bouletreau, J. J. M. van Alphen, J. C. van Lenteren, *Genet. Biol. Drosophila* **3E**, 347 (1986).
26. R. C. King, S. K. Aggarwal, U. Aggarwal, *J. Morphol.* **124**, 143–165 (1968).

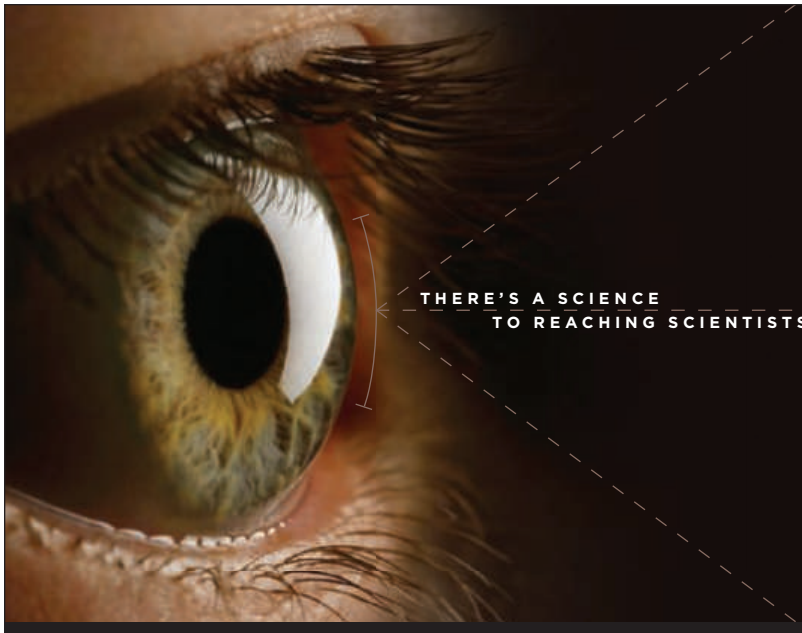
ACKNOWLEDGMENTS

We thank B. Lazzaro for providing bacterial strains and for guidance on experimental design; S. Ruzsa and T. Post for technical assistance; J. Olson and M. Sikes for their assistance with obtaining institutional biosafety approval; J. Mahaffey, J. Alonso, A. Stepanova, and M. Scott for generously providing access to equipment required for this work; and T. F. C. Mackay, E. A. Stone, and D. M. Nielsen for guidance on statistical analysis. Comments from two editors and three anonymous reviewers markedly improved this manuscript. This work was supported by startup funds provided by North Carolina State University to N.D.S. and by NSF grant 1257469 to T.A.S. Data are deposited in the Dryad Repository (doi: 10.5061/dryad.8jj28).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6249/747/suppl/DC1
Materials and Methods
Fig. S1
Tables S1 and S2

22 March 2015; accepted 20 July 2015
10.1126/science.aab1768



Save these Dates!

Postdoc Careers

August 28, 2015

Ads accepted on a first-come, first-served basis

Faculty Careers

September 18, 2015

Reserve ads by September 1 to guarantee space

For recruitment in science, there's only one Science

Two Fantastic Recruiting Opportunities!

POSTDOC CAREERS | August 28, 2015

Be sure to promote your openings to the thousands of scientists who read *Science* to find out about the latest postdoc opportunities.

Special bonus distribution to 25,000 scientists beyond our regular circulation.

Ads accepted on a first-come, first-served basis.

FACULTY CAREERS | September 18, 2015

Gear up to recruit for the faculty positions at your university with this much anticipated issue that reaches thousands of Ph.D. scientists looking for positions in academia.

Special bonus distribution to 25,000 scientists beyond our regular circulation.

Reserve space by September 1, 2015.

Produced by the *Science*/AAAS Custom Publishing Office.

SCIENCECAREERS.ORG

Science Careers

FROM THE JOURNAL SCIENCE AAAS

To book your ad: **advertise@sciencecareers.org**

The Americas
+202-326-6582

Europe/RoW
+44(0)1223-326500

Japan
+81-3-3219-5777

China/Korea/Singapore/Taiwan
+86-186-0082-9345

JOIN AAAS

Get instant access to *Science*. Support all of the sciences.

The American Association for the Advancement of Science (AAAS) is a non-profit community that is open to everyone, from Nobel laureates to high school students. Ours is a global membership of over 120,000 people who believe in the power of science to make the world a better place.

From the moment you join, you get immediate access to everything that AAAS's award-winning journal *Science* has to offer, including:

- 51 weeks of home delivery of *Science*;
- Instant online retrieval of every *Science* article ever published, from today, dating back to 1880;
- Anytime, anywhere access via the *Science* mobile site and apps for Android, iPad, and iPhone devices;
- Members-only newsletters; and more.

As a member, you are also making a critical contribution to AAAS's efforts to provide a public voice for all of science. With public skepticism about science increasing, and public funding for research more uncertain than ever, this work has never been more important.

AAAS is hard at work promoting science in government offices, in schools, and in the public commons all around the world—with programs like AAAS Senior Scientists and Engineers, which brings volunteer scientists into public school classrooms, or our sweeping petition drives calling for the preservation of federal R&D funding.

Visit **promo.aaas.org/joinaaas** and join today. Together we can make a difference.

Science
AAAS

Ambient Ionization Mass Spectrometer

The new AccuTOF-DART4G comes with new performance capabilities including enhanced resolution, speed, and accuracy in a rugged, flexible, versatile design. The AccuTOF-DART 4G couples the facile operation of the DART (Direct Analysis in Real Time) ion source with the high-resolution, accurate mass capability of the AccuTOF time-of-flight mass spectrometer. Not only can the user rapidly acquire data for mixtures and complete unknowns without sample carryover, but also the resulting spectra yield unambiguous assignments, isotopic ratios, and elemental compositions.

Jeol
For info: 978-535-5900
www.jeolusa.com

Lab Management System

The NuGenesis Lab Management System is a unique and powerful alternative to a traditional Laboratory Information Management System with advances that enable deeper insight into scientific challenges, accelerated decision-making, better business results, and compliance with government regulations. The NuGenesis Lab Management System offers high-impact functionality. New advances include NuGenesis SampleShare, an optional, secure web-client for sample submissions and results management; NuGenesis Stability, a complete stability protocol management and testing solution to facilitate a consistent regimented workflow across lab operations; NuGenesis Connectors, a bidirectional link between lab systems and business applications; and Paradigm Scientific Search, a fully integrated scientific search solution for text, documents and science objects. Uniquely combining data, workflow, and sample management capabilities to support the entire product lifecycle from discovery through manufacturing, the NuGenesis Lab Management System is a user-centric platform that encompasses a Scientific Data Management System, Electronic Laboratory Notebook, and Laboratory Execution System.

Waters Corporation
For info: 800-252-4752
www.waters.com/nugenesis

Semi-automatic Electronic Pipette

The CyBi-SELMA is a semi-automatic electronic pipette, equipped with 96 or 384 parallel working tips with a volume range from 500 nL to 1 mL. Based on CyBio's well-proven tip sealing technology, CyBi-SELMA offers error free and reproducible results. Automatic tip tightening avoids laborious adjustments and preloaded CyBi-TipTrays guarantee rapid execution within seconds. An open design allows the processing of any 96- or 384-well microplates as well as the use of accessories like shakers. The pipette features intuitive operation via multilingual touchscreen and a comfortable selection of different pipetting parameters and saving of methods. Thanks to its intuitive touch screen user interface, CyBi-SELMA is amazingly easy to operate. All manual-operating tasks such as tip change or plate change are shown on the display. In addition, any kind of liquid handling parameters like volume and pipetting speed can be adjusted.

Analytik Jena
For info: +49-(0)-36-41-77-70
www.analytik-jena.de

Automated Chemistry Analyzer

The new Flow Solution 3700 Automated Chemistry Analyzer is an advanced continuous flow analyzer with modular hardware that allows FIA, SFA, iSFA, and SFIA methods to be run on the same unit. The FS 3700 incorporates the latest technology that enables higher accuracy and precision of measurement, and the modular design gives the FS 3700 remarkable flexibility and ease of use. The integrated capabilities of the powerful and customizable FlowView software include an intuitive interface, on-the-fly sample addition, and online data processing. A variety of robust detectors, preassembled chemistry cartridges, and plug-and-play valves for FIA/iSFA/SFIA applications offer a new level of performance and versatility. Cartridges for validated USEPA, ASTM, and ISO standardized methods are available, as well as customized cartridges for specialized applications. Accurate, efficient, and modular, the Flow Solution 3700 provides the optimum platform for compliance monitoring, process control, QA/QC, research, and proprietary methods.

Xylem Analytics
For info: 978-778-1010
www.xylemanalytics.com

Molecular Barcoding/Analysis Platform

The GemCode Platform is a unique molecular barcoding and analysis platform that consists of instrumentation, reagents, and software, which delivers long range information, including haplotype phasing, structural variation, and de novo assembly. By upgrading existing short read sequencers the GemCode Platform generates a powerful new data type: linked reads. The GemCode Platform partitions arbitrarily long DNA molecules (including >100 kb) and prepares sequencing libraries in parallel, so that all fragments produced within a partition share a common barcode. A simple workflow combines large partition numbers with a massively diverse barcode library to generate >100,000 barcode containing partitions, while only requiring ~1 ng of DNA input. The GemCode Platform will allow researchers to routinely access genomic information they can't see today.

10X GENOMICS
For info: 925-401-7300
www.10xgenomics.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

THE Welch FOUNDATION

Advancing Chemistry. Improving Life.

Next Generation Medicine

59th Annual Conference on Chemical Research

Joseph L. Goldstein,
University of Texas Southwestern Medical Center

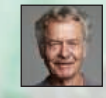
The 2015 Welch Conference on Chemical Research, "Next Generation Medicine," is organized to provide a broad perspective on recent genetic, biochemical, and pharmacological advances that will have a major impact on the future understanding and therapy of human diseases.

An all-star cast of 17 Speakers will present their latest discoveries covering diverse areas of contemporary biomedical science that deal with genome sequencing, mapping the brain, new signaling pathways, molecular pathogenesis of common diseases, and modern approaches to drug therapy.

In addition to the 17 Speakers, 4 prominent scientists will act as Session Leaders who will comment critically on the 17 talks.

2015 Conference on Chemical Research:
Next Generation Medicine
October 26-27 in Houston

Conference complimentary with registration.
Program and registration:
www.welch1.org/chemical-conference



David Botstein, *Calico Life Sciences*

Michael Brown, *University of Texas Southwestern Medical Center*

Zhijian (James) Chen, *University of Texas Southwestern Medical Center*

Titia De Lange, *Rockefeller University*

Karl Deisseroth, *Stanford University*

Hal Dietz, *Johns Hopkins University School of Medicine*

Jennifer Doudna, *University of California, Berkeley*

Helen Hobbs, *University of Texas Southwestern Medical Center*

Tasuku Honjo, *Kyoto University*

Rudolf Jaenisch, *Whitehead Institute for Biomedical Research and MIT*

William Kaelin, *Dana-Farber Cancer Institute, Harvard Medical School*

Eric Lander, *Broad Institute and MIT*

Richard Lifton, *Yale University School of Medicine*

Steven McKnight, *University of Texas Southwestern Medical Center*

Vamsi K. Mootha, *Massachusetts General Hosptial, Harvard Medical School*

Eric Olson, *University of Texas Southwestern Medical Center*

Stanley Prusiner, *University of California, San Francisco*

Gerald Rubin, *Janelia Farm Research Campus, HHMI*

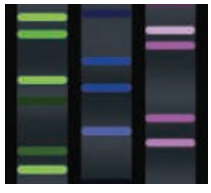
Charles Sawyers, *Memorial Sloan Kettering Cancer Center*

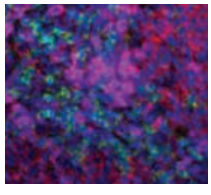
Kevan M. Shokat, *University of California, San Francisco*

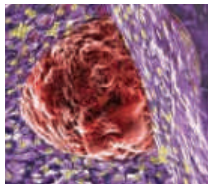
Huda Zoghbi, *Baylor College of Medicine*

want new technologies?

antibodies
apoptosis
biomarkers
cancer
cytometry
data
diseases
DNA
epigenetics
genomics
immunotherapies
medicine
microbiomics
microfluidics
microscopy
neuroscience
proteomics
sequencing
toxicology
transcriptomics







watch our webinars

Learn about the latest breakthroughs, new technologies, and ground-breaking research in a variety of fields. Our expert speakers explain their quality research to you and answer questions submitted by live viewers.

VIEW NOW!
webinar.sciencemag.org

Science
AAAS

Brought to you by the *Science*/AAAS Custom Publishing Office

@SciMagWebinars

Science Careers Advertising

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

Tracy Holmes
Worldwide Associate Director
Science Careers
Phone: +44 (0) 1223 326525

THE AMERICAS

E-mail: advertise@sciencecareers.org

Fax: +1 (202) 289 6742

Tina Burks
Phone: +1 (202) 326 6577

Nancy Toema
Phone: +1 (202) 326 6578

Online Job Posting Questions
Phone: +1 (202) 312 6375

EUROPE / INDIA / AUSTRALIA / NEW ZEALAND / REST OF WORLD

E-mail: ads@science-int.co.uk

Fax: +44 (0) 1223 326532

Sarah Lelarge
Phone: +44 (0) 1223 326527

Kelly Grace
Phone: +44 (0) 1223 326528

Online Job Posting Questions
Phone: +44 (0) 1223 326528

JAPAN

Katsuyoshi Fukamizu (Tokyo)
E-mail: kfukamizu@aaas.org
Phone: +81 3 3219 5777

Hiroyuki Mashiki (Kyoto)
E-mail: hmashiki@aaas.org
Phone: +81 75 823 1109

CHINA / KOREA / SINGAPORE / TAIWAN / THAILAND

Ruolei Wu
Phone: +86 186 0082 9345
E-mail: rwu@aaas.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

POSITIONS OPEN

UNIVERSITY OF MINNESOTA – TWIN CITIES
Biochemistry, Molecular Biology, and Biophysics

The Department of Biochemistry, Molecular Biology, and Biophysics invites applications for a full-time, tenured **ASSOCIATE PROFESSOR** position to begin on or around January 1, 2016. Although all types of biomedical research will be considered, we are particularly interested in candidates with expertise in transcriptional or posttranscriptional gene regulation as well as all aspects of chromatin modification and regulation. The successful candidate will be expected to have a strong, externally funded research program and contribute to the interdisciplinary undergraduate, graduate, and professional teaching programs of the Department. Cross-disciplinary collaboration is also an asset. All candidates must currently be at the Assistant Professor (tenure track) or Associate Professor level with a strong publication record. The successful candidate will receive a startup package and a salary commensurate with education and experience.

The deadline for submitting applications is September 1, 2015. Please apply online at **website: http://umn.edu/ohr/employment** and click the blue "search and apply for current openings" link located in the center of the page. Enter 303175 in the "Keywords" box and click the yellow search button. Please attach full curriculum vitae, a two-page research overview, and a one-page teaching statement. Please attach the names of three references who will be able to provide research and teaching potential.

The University of Minnesota is an Equal Opportunity Educator and Employer. The Department of Biochemistry, Molecular Biology, and Biophysics strongly encourages all individuals to apply regardless of gender, ethnicity, or background and is committed to a diverse research and teaching environment.

TENURE TRACK FACULTY POSITION
Georgia Institute of Technology School of Chemistry and Biochemistry
Atlanta, GA 30332-0400

The Georgia Institute of Technology, School of Chemistry and Biochemistry seeks to fill three or more tenure-track faculty positions. Candidates from all areas will be considered, with opportunities for joint appointments in other departments of science and engineering to facilitate interdisciplinary research and scholarship. Exceptional candidates at all levels are encouraged to apply. Candidates for appointment at the **ASSISTANT PROFESSOR** level should submit an application letter, curriculum vitae, summary of research plans, description of teaching interests and philosophy, and arrange for submission of three letters of reference. Candidates at advanced levels should submit an application letter, curriculum vitae, and a brief description of research plans (particularly if future plans differ significantly from past efforts). All materials and requests for information should be submitted electronically, as per the instructions found at **website: https://academicjobsonline.org/ajo/jobs/5757.**

The application deadline is September 15, 2015 with application review continuing until the positions are filled.

Georgia Tech is an Equal Education/Employment Opportunity Institution.

UNIVERSITY OF MICHIGAN

BIOLOGICAL SCIENCES SCHOLARS PROGRAM
For Junior, Tenure-Track Faculty

The University of Michigan Medical School announces recruitment for the Biological Sciences Scholars Program (BSSP) to enhance the institution's strengths in the biological and biomedical research areas.

Now entering its 17th year, the BSSP has led recruitment of outstanding scientists pursuing research in genetics, microbiology, immunology, virology, structural biology, biochemistry, molecular pharmacology, stem cell biology, cancer biology, physiology, cell and developmental biology, bioinformatics, and the neurosciences. The Program seeks individuals with PhD, MD, or MD/PhD degrees, at least two years of postdoctoral research experience, and who have not previously held a faculty position. Candidates will show evidence of superlative scientific accomplishment and scholarly promise. Successful candidates will be expected to establish a vigorous, externally-funded research program, and to become leaders in departmental and program activities, including teaching at the medical, graduate, and/or undergraduate levels. Primary departmental affiliation(s) will be determined by the applicant's qualifications and by relevance of the applicant's research program to departmental initiatives and themes. All faculty recruited via the BSSP will be appointed at the Assistant Professor level.

APPLICATION INSTRUCTIONS: Please apply to the Scholars Program through the BSSP website at: **http://bssp.med.umich.edu**. A curriculum vitae (including bibliography), a three page research plan, an NIH biosketch, and three original letters of support should all be submitted through the BSSP website. More information about the Scholars Program, instructions for applicants and those submitting letters of recommendation, and how to contact us is located on the BSSP web site: **http://bssp.med.umich.edu**. The deadline for applications is **Friday, October 16, 2015**.

The University of Michigan is an Affirmative Action/Equal Opportunity Employer.

UMBC
AN HONORS UNIVERSITY IN MARYLAND

TENURE TRACK FACULTY POSITION
CHEMICAL BIOLOGY

THE DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY AT THE UNIVERSITY OF MARYLAND, BALTIMORE COUNTY (UMBC) invites applications for a full-time tenure-track faculty position at the Assistant Professor level. Applicants are expected to establish a vigorous and externally funded research program in any sub-discipline of an area *broadly defined as chemical biology* (e.g., bioorganic, bioanalytical, biophysical chemistry, and biochemistry). Candidates must have an outstanding record of scientific achievement, demonstrated by publications in peer-reviewed journals. The successful applicant should have a PhD and postdoctoral experience in any of the sub-disciplines listed above, will be expected to teach at both the undergraduate and graduate (PhD and MS) levels with particular emphasis on bioorganic, biophysical, or biochemistry courses, and should demonstrate an understanding of and commitment to diversity.

The Department (www.umbc.edu/chem-biochem) is a highly cross-disciplinary and interactive group of faculty, postdoctoral fellows, and students engaged in cutting edge research, working in state-of-the-art laboratory facilities in a recently renovated building. The faculty has strong collaborations across the entire campus, and with the University of Maryland Schools of Medicine and Pharmacy.

UMBC is committed to academic excellence and diversity within the faculty, staff, and student body (http://facultydiversity.umbc.edu/). The university is strategically situated on a suburban campus in the intellectually and culturally vibrant Baltimore-Washington corridor, providing unique opportunities afforded by its diversity, intermediate size and world-class infrastructure. Candidates should submit their applications electronically to **http://apply.interfolio.com/30626**, and include curriculum vitae, description of research plans, and statement of teaching philosophy. The candidates also need to arrange for three letters of recommendation to be submitted to **http://apply.interfolio.com/30626**. For inquiries, please email **chemsearch@umbc.edu**. Review of applications will begin **November 1st, 2015** and continue until the position is filled. The appointment will commence in August 2016.

UMBC is an Equal Opportunity/Affirmative Action Employer. Applications from women, minorities, individuals with disabilities and other traditionally under-represented groups in the sciences are especially encouraged.

NYU Langone MEDICAL CENTER

NOBEL PRIZE WINNERS.
WORLD-RENOWNED RESEARCHERS.
AND YOU.

Basic Science & Clinical Research Interview Day

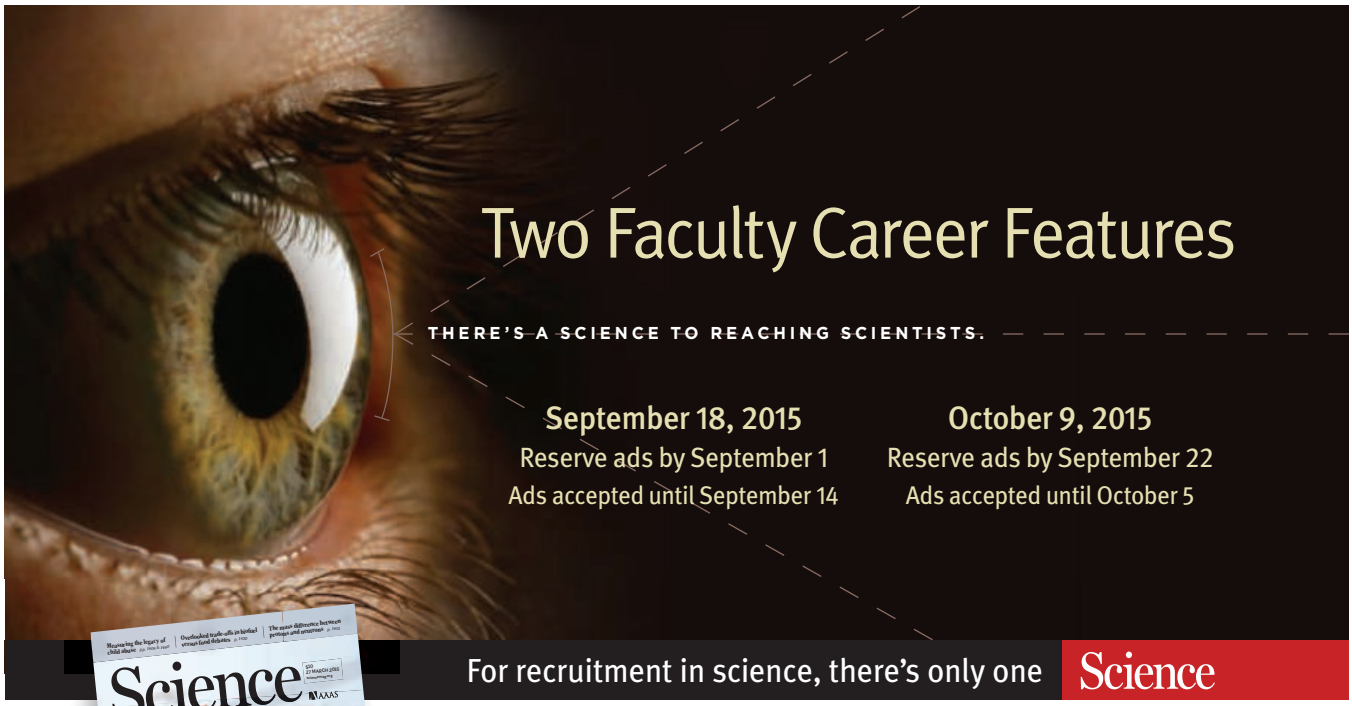
Monday, September 28, 4-8pm

NYU Langone Medical Center has created a stimulating research environment to foster groundbreaking discoveries. We have current funding to support 900 clinical research studies, the technology infrastructure to capture up to 1 petabyte of clinical data, and the intellect to translate our research into medicine. It all adds up to innovative research and extraordinary opportunities.

To explore Scientist, Technician and Research Coordinator careers at NYU Langone, please pre-register for our Interview Day at **careers.nyumc.org/events.**

EOE including Disabled and Vets VEVRAA Federal Contractor

ScienceCareers online @sciencecareers.org



Two Faculty Career Features

THERE'S A SCIENCE TO REACHING SCIENTISTS.

September 18, 2015
Reserve ads by September 1
Ads accepted until September 14

October 9, 2015
Reserve ads by September 22
Ads accepted until October 5

For recruitment in science, there's only one **Science**

Hiring faculty? Whatever your timing, we've got two special features for your **Faculty ads** this fall! The September 18 feature covers strategies and resources to build teaching skills. The October 9 feature offers strategies for moving into academia from other industries.

Special packages available when you advertise in both features

What makes *Science* the best choice for recruiting?

- Read and respected by 570,400 readers around the globe
- 60% of our weekly readers work in academia and 67% are Ph.D.s. *Science* connects you with more scientists in academia than any other publication
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

Why choose these faculty features for your advertisement?

- Relevant ads lead off the career section with special Faculty banner
- September 18 issue will be distributed at the Biotechnica Meeting in Hanover, Germany, 6–8 October.

Expand your exposure. Post your print ad online to benefit from:

- Link on the job board homepage directly to Faculty jobs
- Dedicated landing page for faculty positions
- Additional marketing driving relevant job seekers to the job board.

SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

To book your ad: **advertise@sciencecareers.org**

The Americas
202-326-6582

Europe/RoW
+44(0)1223-326500

Japan
+81-3-3219-5777

China/Korea/Singapore/Taiwan
+86-186-0082-9345

UPMC LIFE CHANGING MEDICINE

University of Pittsburgh
School of Medicine
Chair, Department of Pediatrics
Children's Hospital of Pittsburgh of UPMC

The University of Pittsburgh School of Medicine and Children's Hospital of Pittsburgh of UPMC are seeking applications for the position of Professor and Vira I. Heinz Endowed Chair of the Department of Pediatrics and Physician-in-Chief of the Children's Hospital of Pittsburgh and Scientific Director of the Rangos Research Institute. The department is dedicated to providing high-quality health care while developing and maintaining innovative basic and clinical research programs aimed at advancing clinical practice. The department has approximately $29 million per year of federal funding and $39 million of total extramural funding. It supports a large and outstanding residency program with multiple tracks, as well as 17 fellowship programs, 6 NIH T32 grants for postdoctoral research training, and a K12 training grant for junior faculty researchers.

Competitive candidates must have academic accomplishments meeting criteria for appointment at the tenured professor level, including a distinguished record of research, clinical, teaching, and service activities. Other key characteristics include a broad visionary approach to major issues in the field, ability to foster collaborations, capability to manage an active group of community and academic specialty practices, as well as significant administrative and leadership experience.

The University of Pittsburgh School of Medicine is one of the nation's leading medical schools, renowned for its curriculum that emphasizes both the science and humanity of medicine and its remarkable growth in NIH funding. With more than $450 million of NIH funding, the University ranks fifth among more than 3,000 entities that receive NIH support. As one of the University's six Schools of the Health Sciences, the School of Medicine is the academic partner to the University of Pittsburgh Medical Center. The combined mission is to train tomorrow's health care specialists and biomedical scientists, engage in groundbreaking research that will advance understanding of the causes and treatments of disease, and participate in delivery of outstanding patient care. UPMC is a global health system with 23 hospitals, several international sites, more than 60,000 employees, and close to $13 billion of annual revenue.

Review of applications will begin immediately and continue until the position is filled. Candidates should submit a letter of application stating professional accomplishments; curriculum vitae; and the names, mailing addresses, e-mail addresses, and telephone numbers of five professional references. Electronic applications and nominations are preferred and should be sent to **mayowski@pitt.edu** with **"Pediatrics Chair Search"** entered into the e-mail subject line. Applications and nominations may also be submitted by mail and should be sent to: **Ann E. Thompson, MD, Vice Dean, Attention: Colleen Mayowski, EdD, University of Pittsburgh School of Medicine, M240 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261; 412-648-9000; E-mail: mayowski@pitt.edu**

The University of Pittsburgh and UPMC are Affirmative Action, Equal Opportunity Employers.
Women and minority candidates are strongly encouraged to apply.

Science Careers
Cernet

"《科学》职业" 已经与Cernet/赛尔互联开展合作。中国大陆的高校可以直接联系Cernet/赛尔互联进行国际人才招聘。

请访问 Sciencecareers.org/CER 点得联系信息。

中国大陆高校以外的 招聘广告，或者高校的其它业务，
请与国际合作、出版副总监吴若蕾联系：
+86-186 0082 9345 rwu@aaas.org

招募学术精英，《科学》是您的不二之选 Science

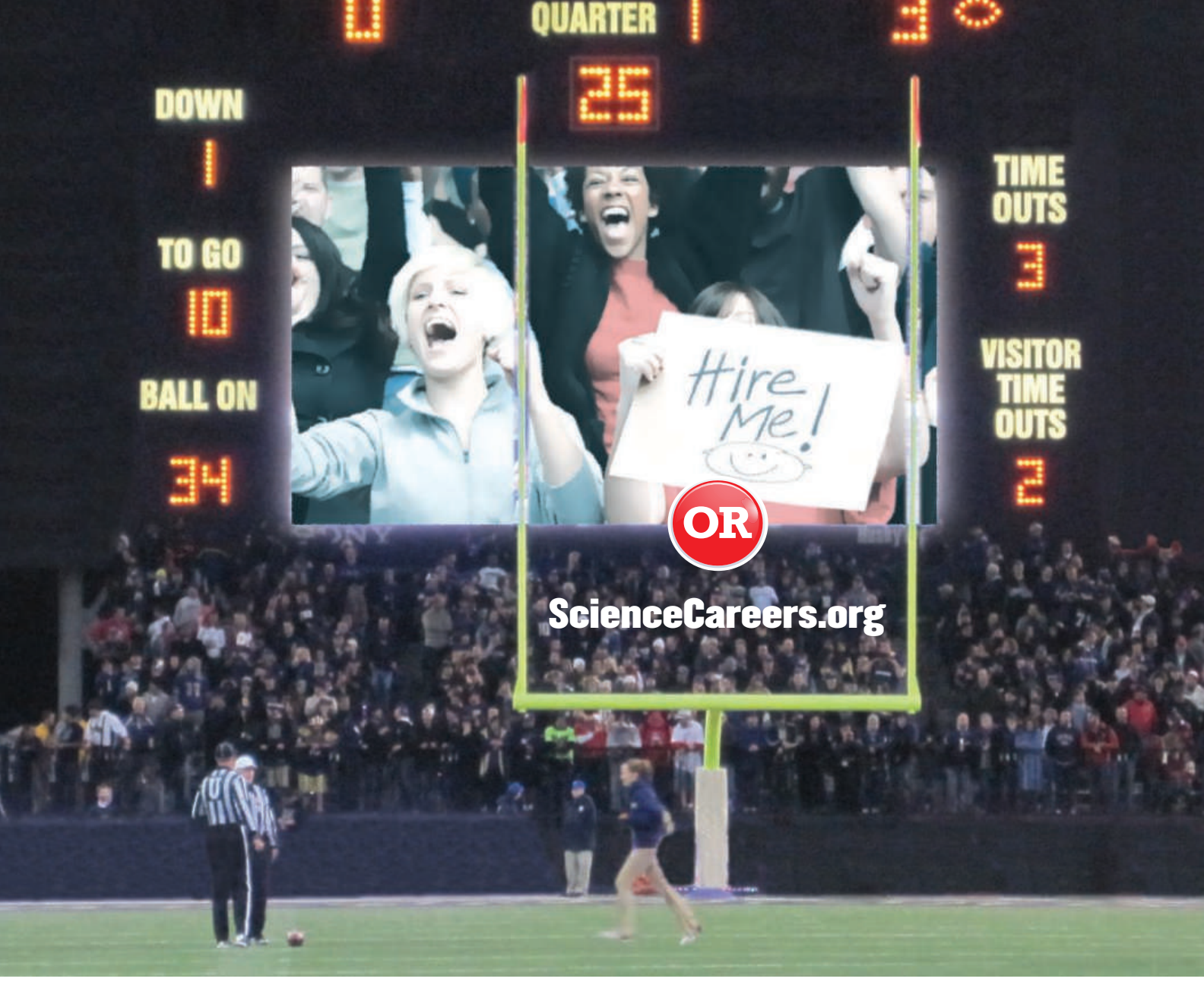
UNIVERSITY OF MICHIGAN

Ecology or Evolutionary Biology of Fishes or Birds

The Department of Ecology and Evolutionary Biology (www.lsa.umich.edu/eeb) and the Program in the Environment (www.lsa.umich.edu/pite) at the University of Michigan seek applicants for an assistant professor (tenure-track) position in the ecology or evolutionary biology of fishes or birds. While we expect to make a junior hire, outstanding senior applicants will also be considered. This is a university-year appointment with an expected start date of September 1, 2016. We seek outstanding individuals who use comparative fish or bird systems to study any area of ecology or evolutionary biology, and who would offer exceptional courses in the ecology or evolution of either taxon. Also strongly encouraged are research programs that could take advantage of the world-class biodiversity collections of the Museum of Zoology and/ or utilize the EEB Department's biological field stations. Museum curatorial activities may replace some teaching duties for appropriate candidates.

To apply, use this link – **http://www.resources-eeb.lsa.umich.edu/search15** - and arrange to have three letters of recommendation submitted through the same website. Review of applications will begin on **October 1st 2015** and will continue until the position is filled.

Women and minorities are strongly encouraged to apply. The University of Michigan is supportive of the needs of dual career couples and is an Equal Opportunity/Affirmative Action Employer.



Learn more and keep your job search out of the cheap seats.

- Search thousands of job postings
- Create job alerts based on your criteria
- Get career advice from our Career Forum experts
- Download career advice articles and webinars
- Complete an individual development plan at "myIDP"

Target your job search using relevant resources on **ScienceCareers.org**.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

myIDP: A career plan customized for you, by you.

For your career in science, there's only one Science

my IDP
ScienceCareers

Recommended by leading professional societies and the NIH

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

advancing the life sciences 1912-2012 FASEB Federation of American Societies for Experimental Biology

UCSF University of California San Francisco

Medical College of Wisconsin

Burroughs Wellcome Fund

ScienceCareers online @sciencecareers.org

By **Jeffrey J. McDonnell**

Creating a research brand

In my first few years beavering away on the tenure track, I flitted from one research question to another. In an undisciplined pursuit of funding, I wasted my energies—and, sadly, papers—on unrelated topics. I began to understand how a successful research program worked only after I became an associate professor at another university. I realized from observing others that it was critical to find one's focus and voice and have it heard—quickly. Now, when my own students and postdocs ask me how they can stand out among the many young scientists jockeying for position and positions, I tell them to learn from my early struggles, and to define their research brand identity.

A research brand identity, in the best scientific sense, is your central mission: the particular branch of research you'd like to be truly excellent in and known for. Equal parts research focus and career vision, a brand identity is a map of a few key subareas where you strive to make a difference over a sustained period. By crafting a research brand, you make it clear to others how they should define you, and you help them remember who you are and what you do.

As a young hydrologist, I worked on topics as varied as soil erosion, snow physics, and the remote sensing of water uptake by plants. When I finally found my focus, I concentrated on three simple questions: Where does water go when it rains? What pathway does it take to the stream? How old is the water in our streams? At first these questions were narrow; I explored just beyond the boundaries of my thesis work, in a space where I had established some credibility. Over time the questions broadened. I now explore where plants get water and the role of geology in defining water storage and release. These core issues have sustained me over my career.

The trick is to find the optimal research brand width: too broad and you blend in with others and lack a niche to carve out; too narrow and you seem too limited. You want some connection to your adviser's research brand, but you also want to move in a new direction that honors your unique skills. The old adage "only do what only you can do" can be helpful as you ponder this.

After you have identified your brand identity, getting it out into the world is relatively straightforward: Get good results, publish papers, and give talks that all map to your theme. But there is much more. The first step is describing your brand succinctly on your research homepage. The second, and more active step, is promoting it through a set of activities directed at your peers. You might write a journal commentary, perspective paper, or review linked to your brand. You might guest edit a journal special issue or edit a research text. Or you could organize a conference session or a boutique research conference focused on your theme. The idea is not to do just one of these things but to do as many as you can.

"Get good results, publish papers, and give talks that all map to your theme."

All of these activities build a whole that is greater than the sum of the parts. Come tenure time, you want to be known for something, and you want external reviewers examining your dossier to know of you. This strategy will almost guarantee that.

Finally, a research brand identity remains helpful as you advance. As more opportunities come your way, your brand can help you filter them, keeping your main thing the main thing. The brand helps you determine what to say no to, a skill that becomes much more important as you grow more sought after over time. And just as a research brand identity is a way to describe your ambitions when you start out, it is also a way to sum up your body of work later in your career. ■

Jeffrey J. McDonnell is a professor in the School of Environment and Sustainability at the University of Saskatchewan in Saskatoon, Canada, and Sixth Century Chair at the University of Aberdeen in the United Kingdom. *He is grateful to Maureen Reed for ongoing discussions and to his lab group for useful feedback.* *For more on life and careers, visit sciencecareers.org. Send your story to SciCareerEditor@aaas.org.*

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS